

# Spatiotemporal modelling in biology:

from transcriptional regulation to  
plasmid positioning

JALDERT HUGO RIGOBERT IETSWAART

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia  
John Innes Centre  
September, 2015

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.



# ABSTRACT

---

Here I describe how cycles of mathematical modelling and experimenting have advanced our quantitative understanding of two different processes: transcriptional regulation of the floral repressor *FLOWERING LOCUS C* (*FLC*) in *Arabidopsis thaliana* and spatial positioning of low copy number plasmids in *Escherichia coli*. Despite the diversity in biological subjects, my spatiotemporal modelling approach provides a common ground.

*FLC* regulation involves an antisense-mediated chromatin silencing mechanism, where alternative polyadenylation of antisense transcripts is linked to changed histone modifications at the locus and altered expression. Mathematical model predictions of *FLC* transcriptional dynamics are validated by measurements of total and chromatin-bound *FLC* intronic RNA. This demonstrates that *FLC* regulation involves a quantitative coordination between transcription initiation and elongation, potentially a general feature of gene regulation in a chromatin context. A quantitative analysis of cellular RNA levels indicates that *FLC* processing and degradation are well described by Poisson processes. *FLC* transcription correlates with cell volume, which underlies the large cellular variation in transcript levels.

Low copy number plasmids in bacteria require segregation for stable inheritance through cell division. This is often achieved by a *parABC* locus, comprising an ATPase ParA, DNA-binding protein ParB and a *parC* region, encoding ParB-binding sites. These components space plasmids equally over the nucleoid, yet the underlying mechanism has not been understood. Here I show mathematically that differences between competing ParA concentrations on either side of a plasmid can specify regular plasmid positioning. This can be achieved regardless of the exact mechanism of plasmid movement. Experimentally, *parABC* from *E. coli* plasmid pB171 increases plasmid mobility, inconsistent with models based on plasmid diffusion and immobilization. Instead this observation favours a directed motion model. These results unify previously contradictory models for plasmid segregation and provide a mechanistic basis for self-organized plasmid spacing.





# CONTENTS

---

ABSTRACT 3

LIST OF PUBLICATIONS 9

PREFACE 11

ACKNOWLEDGEMENTS 13

1 INTRODUCTION 15

1.1 Flowering time control: on the connection between antisense RNA and chromatin 15

1.1.1 Interplay between transcription and chromatin 15

1.1.2 Widespread antisense transcription 21

1.1.3 Antisense-mediated regulation in plants 23

1.1.4 Introduction to *FLC* regulation 25

1.1.5 Autonomous pathway 26

1.1.6 Vernalization 29

1.1.7 Genetic toggle mechanism 33

1.1.8 RNA turnover linked to histone modifications 33

1.1.9 Antisense RNA in Polycomb silencing 36

1.2 Regular positioning of low copy number plasmids 37

1.2.1 DNA segregation in prokaryotes 37

1.2.2 The ParA family of plasmid partitioning mechanisms 39

1.2.3 Regular positioning of other subcellular machinery 43

1.2.4 Cellular architecture of prokaryotes 43

1.3 Concluding remarks 45

2 COORDINATION OF *FLC* TRANSCRIPTION INITIATION AND ELONGATION 47

2.1 Introduction 47

2.2 Results 48

2.2.1 *FLC* expression changes are consistent with regulation of transcription 48

2.2.2 Effects of initiation, elongation and termination on Pol II levels 51

2.2.3 RNA fold changes do not reflect Pol II occupancy changes 53

2.2.4 *FLC* transcriptional dynamics can be explained by coordination of initiation and elongation. 54

2.2.5 Initiation and elongation rate effects on intronic RNA expression 62

2.2.6 Coordinated initiation and elongation generates RNA upregulation patterns along *FLC* intron1. 65

2.2.7 Sense premature termination contributes little to *FLC* repression 69

2.2.8 Co-transcriptional 5' to 3' degradation contributes little to *FLC* repression 74

2.2.9	<i>FLC</i> transcriptional repression involves local FLD activity and an altered chromatin state	79
2.2.10	Stochastic model based on FLD action through FCA-dependent proximal <i>COOLAIR</i> processing	79
2.2.11	Stochastic coordination model predicts that slow Pol II elongation is required for <i>FLC</i> repression	92
2.3	Discussion	93
2.4	Materials and Methods	96
2.4.1	Determining the Pol II ChIP fragment size distribution	96
2.4.2	Convolution of the Pol II density for comparison with experiments	96
2.4.3	Parameter estimation of the co-transcriptional degradation model	99
3	VARIATION IN <i>FLC</i> TRANSCRIPTION, PROCESSING AND DEGRADATION	105
3.1	Introduction	105
3.2	Results	106
3.2.1	Cellular <i>FLC</i> distribution is monomodal and super-Poissonian	106
3.2.2	<i>FLC</i> degradation satisfies Poisson characteristics	110
3.2.3	A constant transcription rate per locus is not sufficient to explain <i>FLC</i> variation	113
3.2.4	A transcription rate scaling with cell volume can explain <i>FLC</i> variation	116
3.2.5	<i>FLC</i> RNA accumulates little at the locus	118
3.2.6	Estimation of <i>FLC</i> and <i>COOLAIR</i> production and splicing rates	121
3.2.7	Estimation of Pol II elongation rates	123
3.2.8	<i>FRI</i> -mediated coordination of transcription initiation and elongation	124
3.2.9	Stochastic coordination model predicts that fast Pol II elongation enables a high <i>FLC</i> expression state	127
3.3	Discussion	133
3.4	Materials and Methods	135
3.4.1	Probability distributions arising from birth-death processes	135

3.4.2	Model with constant production rate and varying <i>FLC</i> copy number	140
4	EQUAL SPACING OF BACTERIAL PLASMIDS	143
4.1	Introduction	143
4.2	Results	144
4.2.1	ParB-GFP foci are spaced equally over the nucleoid.	144
4.2.2	Mathematical analysis shows that dynamic, competitive ParA concentrations can generate equal plasmid spacing.	146
4.2.3	Diffusion/immobilization model could space highly mobile plasmids equally over the nucleoid.	151
4.2.4	Free plasmid mobility is too low for a diffusion/immobilization mechanism.	155
4.2.5	ParA structures competing to direct plasmid motion can space plasmids equally over the nucleoid.	162
4.2.6	ParA-GFP oscillations are not continuously required for equal plasmid spacing.	167
4.2.7	ParA-GFP forms structures within the nucleoid region.	172
4.2.8	Equal plasmid spacing is compromised in cells with a perturbed nucleoid.	174
4.3	Discussion	182
4.4	Materials and Methods	187
4.4.1	Mathematical derivation that dynamic ParA concentrations can generate equal plasmid spacing.	187
4.4.2	Total ParA fluorescence analysis	192
4.4.3	Total ParB fluorescence analysis	192
4.4.4	Diffusion/Immobilization model	193
4.4.5	Directed motion model	195
4.4.6	Three dimensional nucleoid and ParA structure analysis	196
5	CONCLUSIONS & DISCUSSION	199
5.1	Autonomous pathway regulation of <i>FLC</i>	199
5.2	Cellular variation of <i>FLC</i> expression	201
5.3	Equal plasmid spacing over the nucleoid	202
5.4	Mathematical modelling methodology	204
5.5	Outlook	205
	BIBLIOGRAPHY	207
	APPENDIX: PUBLICATIONS ASSOCIATED TO THIS THESIS	229



# LIST OF PUBLICATIONS

---

This thesis includes material from the works below.

\* Indicates that these authors contributed equally.

R. Ietswaart, Z. Wu, C. Dean. **Flowering time control: another window to the connection between antisense RNA and chromatin.** *Trends Genet.* 28: 445-53 (2012)

R. Ietswaart\*, F. Szardenings\*, K. Gerdes, M. Howard. **Competing ParA structures space bacterial plasmids equally over the nucleoid.** *PLoS Comput Biol.* 10(12): e1004009 (2014)

Z. Wu\*, R. Ietswaart\*, F. Liu, H. Yang, M. Howard and C. Dean. **Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation.** *Proc Natl Acad Sci U S A.* 113(1): 218-23 (2016)

R. Ietswaart\*, S. Rosa\*, Z. Wu\*, M. Hartley, C. Dean and M. Howard. **Quantification of *Arabidopsis FLC* transcription initiation, elongation, splicing, export and degradation dynamics.** *in preparation*



# PREFACE

---

This doctoral thesis is the result of scientific research that I have performed in the past four years. It consists broadly of two biological topics: *FLC* transcriptional regulation in *Arabidopsis* and spatial plasmid positioning in *E. coli*. The commonality between these two lines of research lies in the mathematical modelling that I have conducted. This is reflected in the described results and the style of writing. The research has been highly interdisciplinary and performed in close collaboration with several experimental biologists. Therefore this thesis is written from the first person plural (we) perspective. This thesis also describes several sets of experimental data that I have used to justify my conclusions. Wherever experimental data is described, I acknowledge the contributions of the colleague(s) who generated the data. Methodology on experimental procedures of the respective experiments falls beyond the scope of this thesis. Relevant data analyses procedures that were developed by myself are described.

The targeted readership are mathematically proficient scientists, who have an undergraduate-level understanding of molecular and cellular biology. The biology required to understand the relevance and context of the research will be introduced and in that respect this thesis is self-contained. I assume the reader has sufficient understanding of the required mathematical techniques without further introduction. Chapter 1 is an introduction to the biology of transcriptional gene regulation in general and specific to *FLC*, with my primary research results on the topic described in chapters 2 and 3. Chapter 1 also contains a brief introduction to the field of bacterial plasmid positioning, with subsequent primary research results following in Chapter 4. Chapter 5 constitutes a conclusion and discussion of the research described in this thesis.





## ACKNOWLEDGEMENTS

---

I would like to thank my primary supervisor Martin Howard, who has been very supportive, always reasonable, and fair. As a pioneer of modelling in biology he has taught me many valuable lessons. I appreciate how he has mentored me in difficult situations during my PhD. Thank you also to my secondary supervisor Caroline Dean, a scientific and professional leader with a remarkable biological intuition. Martin and Caroline have both formed a deep source of scientific inspiration for me. They have helped me develop from a student into a scientist. I thank Zhe Wu for the close and sometimes intense collaboration on the autonomous pathway throughout my PhD. Without our numerous scientific discussions and his experiments we could not have made the discoveries on *FLC* coordination of initiation and elongation. Also my appreciation to Florian Szardenings for making our long distance collaboration become fruitful. Without his experiments, we could not have made the discoveries on plasmid positioning. Thank you to Stefanie Rosa and Susan Duncan for their interest in collaborating with me to quantitatively investigate the cellular variation in *FLC* expression. To Matthew Hartley and Tjelvar Olsson for their scientific computing support. Andrew Angel, David Richards, Antonio Scialdone, Scott Berry, Seán Murray, Rea Kourounioti, Giuseppe Facchetti and Leonie Ringrose for scientific discussions and creating a vibrant, friendly atmosphere in the Howard group. In addition to those mentioned above, thanks also to all other current and previous Dean lab members that shared their expertise on *FLC* regulation, in particular Jie Song, Oleg Rait-skin, Pedro Crevillén, Zhiwei Wang, Fuquan Liu, Ryo Ishikawa, Cagla Sonmez, Peijin Li, Danling Zhu, Hongchun Yang and Julia Questa. Nick Pullen for providing his excellent  $\LaTeX$  typesettings from his PhD thesis. The other PhD students and postdocs in the Computational and Systems Biology department at the John Innes Centre (JIC) for sharing interest in mathematical and computational approaches in biology. More broadly the scientific community of JIC, a great environment for plant and microbial science, which is reflected in the topics of my research. The anonymous reviewers of my manuscripts

who have made me realize that scientific progress is tightly connected to human interest. Lastly, the staff members at Santander Corporate and Commercial across East Anglia, in particular Mike Hanby and Tracy Grazioli, for enabling me to be part of their work force during my internship there.

On a more personal note, I would like to thank the many friends that I have met in Norwich, notably some of my year mates who made my PhD life at JIC more relaxed and fun: John, Richard, Mike, Annis, Claire, Jo, Rachel and Chris. My (previous) house mates Kate-rina, Lewis, Ola, Josué, Anil and Andrzej. Jon for developing our side project. Also a big thanks to the Natuurko's: Keesjan, Benjamin, Bas v. O., Chris, Bas C., Ralph and Anton, for their friendship and shared interest in a quantitative approach to science and life! Ricardo, for our long lasting close friendship, and for discovering Middle America together, a truly enriching experience that has changed my attitude towards life. My PhD has resulted in many periods of solitude and focus on work. I am most grateful to have met Ly during this busy part of my life. Her kindness, enthusiasm and endless positivity towards life have helped me stay motivated and hopeful in tough times.

Laurien en Siep, ik wil jullie bedanken voor jullie liefde en steun, voor de vele jaren van opvoeding en investering in mij, en voor jullie geloof in mijn kunnen. Op de belangrijke moment herinnerden jullie mij eraan om me te blijven richten op de juiste zaken. Daarnaast hebben jullie me geholpen om een balans te behouden tussen werk en leven. Ondanks de fysieke afstand zijn jullie altijd dichtbij me gebleven.

Robert Ietswaart  
JOHN INNES CENTRE, NORWICH  
September 2015

# INTRODUCTION

---

In this chapter we first introduce relevant concepts of transcription, chromatin and antisense RNA (asRNA) regulation in plants and other eukaryotes that are relevant to the primary research chapters 2 and 3. We begin by a general introduction on the interplay between transcription and chromatin and proceed with examples where the mechanism of antisense transcription has been analysed and shown to play an important role in chromatin regulation. We focus on the topic of our study, the regulation of the floral repressor *FLOWERING LOCUS C (FLC)* in *Arabidopsis thaliana*, and then review specific examples in *S. cerevisiae* (yeast) and mammals in order to highlight conceptual parallels.

In the second part of this chapter we introduce the topic of bacterial low copy number plasmid positioning. We begin by reviewing the components involved in this process and in other prokaryotic DNA segregation mechanisms. We then consider regular positioning of other sub-cellular components and discuss these systems in the context of the spatial architecture of the cytoplasm in prokaryotes. The relevant primary research on spatial positioning of low copy number plasmids is then described in chapter 4.

## 1.1 Flowering time control: on the connection between antisense RNA and chromatin

### 1.1.1 *Interplay between transcription and chromatin*

Transcriptional regulation is central for changing gene expression patterns that underpin growth and development in all organisms. Many components of the transcriptional machinery are conserved between organisms [1], further indicating their fundamental importance to life. One example of such a core component is RNA Polymerase, a protein complex that transcribes DNA in prokaryotes. In eukaryotes, many species have multiple RNA Polymerase complexes, often with divergent roles in gene regulation [1, 2]. For instance in

*Arabidopsis* there are five different complexes [3]: some function in ribosomal RNA production and others in small interfering RNA production [1–3]. RNA Polymerase II (Pol II) functions to transcribe most protein coding genes, not only in plants but also all other eukaryotes [1].

In eukaryotes, DNA is wrapped around histones to form a unit called a nucleosome [1]. To condense DNA into a size that fits in the nucleus, many nucleosomes are packed together to form chromatin [1]. The nucleus is further organized into higher order structures, such as topologically associating domains (TADs) in metazoans. TADs are formed by DNA sequence elements that are in close physical proximity, despite sometimes being far apart sequencewise in the genome [4, 5]. Transcription regulation is impacted by the many layers of organization in the nucleus. Genes common to a TAD often correlate in gene expression [4]. Moreover, enhancers, discrete DNA elements (50-1500kbp) that contain specific sequence motifs, have been found to regulate transcription by Pol II at distances as large as a million basepairs [6, 7].

Crucially, transcription occurs in a chromatin context. In recent years, our understanding of this connection between chromatin and transcription has made significant progress and it is becoming clear that chromatin modifiers play an integral part in transcriptional regulation [7]. The chromatin structure is regulated by chromatin modifiers through nucleosome positioning and posttranslational modifications of histone tails (Figure 1.1), e.g. histone H3 lysine 4 trimethylation (H3K4me3) [1]. Also transcription leads to recruitment of specific chromatin modifiers that affect the chromatin state [7]. It is still debated whether and how these modifiers and chromatin states cause altered transcription or only result from altered transcription regulation by, for instance, transcription factors (TFs) [7, 8]. Here we introduce some of the many features of transcription by Pol II that are known to be related to *FLC* regulation.

Transcription by Pol II involves three keys steps: initiation, elongation and termination, each of which can be affected by chromatin

structure [9, 10]. For example, *in vitro* reconstituted heterochromatin can inhibit Pol II initiation and elongation [11]. Transcriptional activation often involves DNA sequence specific transcription factors (TFs) that can bind to the gene promoter [1]. The nucleosomal arrangement of DNA wrapped around histones can act as a barrier to Pol II binding [9, 12]. To overcome this hurdle, TFs can interact with cofactors, protein complexes without DNA binding properties, that help activate or repress a gene [7]. One such cofactor, the SWI/SNF chromatin remodelling complex [13], can reposition nucleosomes to facilitate recruitment of Pol II to form a pre-initiation complex (PIC) comprised of many general transcription factors [14]. Once the PIC is formed, Pol II is then phosphorylated at the Serine 5 residue (S5P) at its C-terminal domain [15–17]. This S5P-CTD interacts with enzymes responsible for 5' RNA capping [18], a process that prevents nascent RNA 5'-3' degradation [1].

The Pol II S5P-CTD domain also recruits a histone modifying complex, Set1/COMPASS (Figure 1.1), to trimethylate histone H3 lysine 4 (H3K4me3) [20]. Consistently, H3K4me3 is genome-wide most abundant near gene 5' ends and correlates well with transcriptional activity [21]. H3K4me3 deposition is also dependent on histone H2B lysine 123 (H2BK123) ubiquitination by the RAD6-BRE1 complex [22, 23]. H3K4me3 itself seems to affect transcription indirectly through its interaction with a panoply of factors [24], including the Pol II associated factor 1 Complex (Paf1C) and FACT, a protein complex that allows Pol II to traverse nucleosomes [25]. This modification also affects RNA splicing as it indirectly recruits spliceosome components [25]. Overall, H3K4me3 is widely regarded as a histone modification indicative of active transcription [24].

The effects of H3K4 dimethylation (H3K4me2), deposited co-transcriptionally throughout genes by Set1, seems less clear cut in *Arabidopsis* [26] and often dependent on the context of overlapping non-coding transcription in yeast [27, 28]. This modification recruits a histone deacetylation complex (HDAC), which then represses initiation locally [29]. Through 3'-end antisense transcription, sense promoter located H3K4me2 can then suppress sense transcription through these HDACs [27].

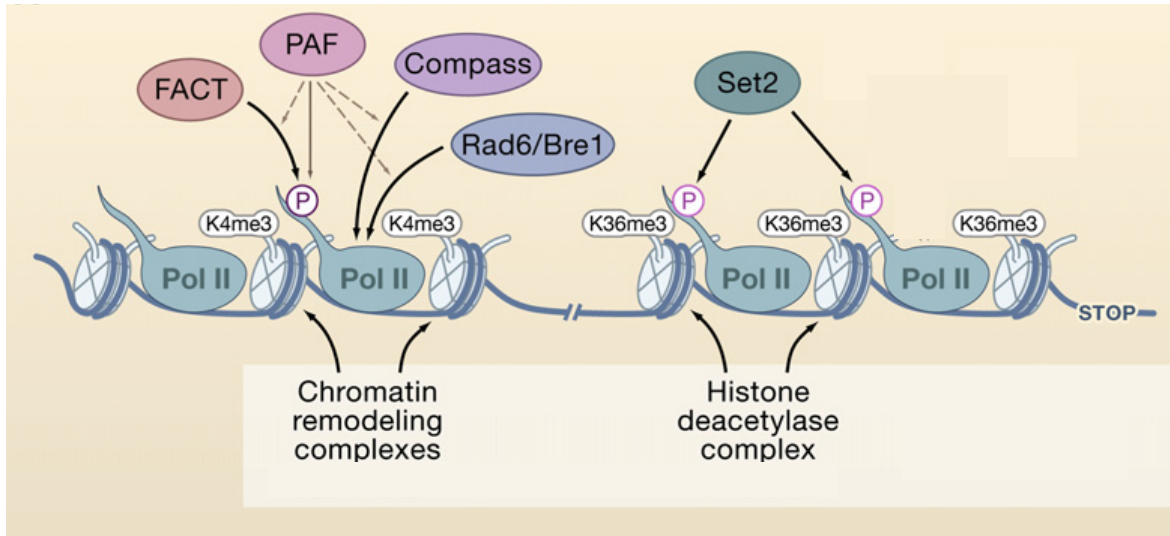


Figure 1.1: Transcription occurs in a chromatin context. The chromatin landscape during transcription elongation is determined by the factors associated with different forms of Pol II. Paf1C (PAF) facilitates the binding of FACT, COMPASS, and Rad6/Bre1 to the S5P-CTD, which results in H2B ubiquitination and accumulation of trimethylation of H3K4 at the 5' end of the gene. Set2 directly interacts with S2P-CTD, thus methylating H3K36 at the 3' end. It is still debated whether and how these modifiers and chromatin states cause altered transcription or only result from altered transcription regulation. This illustration is reprinted from [19] with permission from Elsevier.

However H3K4me2-dependent sense upregulation is also observed for some genes depending on the non-coding gene conformations [28].

Another histone modification, Histone H3 lysine 36 trimethylation (H3K36me3) was first found to be placed co-transcriptionally in yeast by Set2 (Figure 1.1) [30, 31]. Like H3K4me2, it functions to prevent spurious intragenic initiation through recruitment of HDACs [30, 31] and to prevent trans-histone exchange during Pol II elongation [32]. Consistently, H3K36me3 levels are enriched over genic regions genome-wide in many organisms [19, 33]. However more recently, this modification is also been implicated in the developmental regulation of gene expression programs in higher eukaryotes [7, 34]. For instance *C. elegans*, H3K36 trimethylation can also be deposited independently of transcription in order to transmit memory of germline expression to progeny [34]. Overall, as for H3K4me3, H3K36me3 is widely regarded to be a histone modification associated with active transcription [7].

After Pol II initiation and S5 phosphorylation, Pol II proceeds with elongation through the gene in yeast [17]. The Pol II CTD S2 residue is phosphorylated by P-TEFb, a kinase containing protein complex, during elongation within the first 600-1000 base pairs [17, 35]. Pol II can pause and back track throughout elongation [36]. In *Drosophila* however, S5P-Pol II pauses predominantly at a promoter proximal site prior to S2 phosphorylation [37]. After S2 phosphorylation, Pol II is released into elongation, although a minor Pol II fraction terminates prematurely at the promoter proximal pause site [38, 39]. During elongation Pol II is generally very processive *in vitro* and *in vivo* due to the high stability of the RNA-DNA-Pol II ternary complex [1, 9]. In human cells both histones and DNA-bound TFs can affect Pol II elongation locally *in vivo* [40].

During elongation, Pol II generates nascent RNA, which can then be further processed [1]. One such form of RNA processing is splicing: the removal of intron sequences (introns) from newly transcribed RNAs by the spliceosome [1]. The DNA sequences that are not re-

moved from the RNA are termed exons. When the gene encodes for a protein, the exonic RNA parts generally constitute a messenger RNA (mRNA) that can be translated in the cytoplasm [1]. Similar to RNAs, intron boundaries are described by a 5' end and a 3' end, also referred to as the intron donor site and intron acceptor site respectively [1]. During splicing, first the intron donor site and then the acceptor site are cleaved by the spliceosome, after which the remaining exons are ligated together. Splicing can only initiate when Pol II has transcribed the full intron, because 5' cleavage only occurs during lariat formation, a process where the intron 5' and 3' end are connected to form a looped intron RNA (lariat RNA) [1]. Splicing usually occurs at a gene locus, either co-transcriptionally or after Pol II elongation [41]. As alluded to above, splicing is also connected with transcription and chromatin. For instance Pol II can transiently pause at exon 5' ends, presumably to facilitate co-transcriptional RNA splicing [40]. Splicing also feeds back to the chromatin state, as splicing inhibition impairs recruitment of the human H3K36 methyltransferase Setd2 [42]. Alternative splicing and Pol II elongation can be kinetically coupled [43, 44]. In this mechanism splicing of a proximal intron can occur in a 'window of opportunity' during Pol II elongation. Fast Pol II elongation would then lead to a greater chance to by-pass this window, resulting in splicing at a distal 3' splice site.

The process of Pol II termination, also dependent on chromatin remodeller activity [45], generally commences when Pol II transcribes an AATAAA consensus sequence (denoted as a poly(A) site) [1, 46]. The 3' processing factor complexes CPSF and CstF, that associate with Pol II throughout elongation, can recognize this sequence during elongation [47]. In this canonical scenario, Pol II pauses as a result of poly(A) site recognition and both CPSF and CstF promote endoribonucleolytic cleavage [47]. A poly(A) polymerase then adds a poly(A) tail to the RNA 3' end [47]. Pol II termination occurs by the 'torpedo' mechanism [48, 49], where the 5' end of the nascent Pol II-bound RNA is recognized by Xrn2, a conserved chromatin associated 5'-3' exoribonuclease [50]. Xrn2 degrades this RNA until it reaches the Pol II which then drops off the DNA template [48,



49]. More recently, Pol II has also been found to terminate under the influence of cotranscriptional cleavage (CoTC)-type terminator elements [51]. In this case Pol II elongates past the poly(A) site and a downstream CoTC element promotes cleavage and subsequent Pol II termination via the 'torpedo' mechanism. CoTC can depend on various factors, for instance a Dicer endoribonuclease [52], or RNA helicase Sen1[53]. Gene expression can also be regulated through alternative polyadenylation and Pol II termination [54–56]. Premature termination can be kinetically coupled to Pol II elongation [53]. RNA decapping also functions to stimulate premature termination through the 'torpedo' mechanism [54, 57]. Splicing seems to antagonize premature termination, as functional disruption of the spliceosome increases premature termination [58]. Differential splice and poly(A) site signals thereby limit pervasive transcription throughout the genome in humans [58]. Chromatin assembly also limits pervasive non-coding transcription in yeast [59].

RNA 3' processing and export into the cytoplasm generally follows after Pol II termination [1, 41]. RNA export machinery recruitment occurs co-transcriptionally in yeast whilst in humans RNA splicing contributes to this recruitment [60]. Furthermore, aberrant pre-mRNA can be degraded co-transcriptionally upon inhibition of splicing [61]. RNA degradation can occur at many stages during RNA maturation [62] and it seems that in the nucleus the fate of transcripts is largely determined during 3' end formation [63, 64].

As detailed above, protein coding gene regulation already comprises many layers of complexity in a chromatin context. Moreover, antisense RNA has been found to be involved in transcription and chromatin regulation. Below we elaborate on its connection to chromatin and transcription generally in plants and other eukaryotes, and more specifically at *FLC*.

#### 1.1.2 *Widespread antisense transcription*

The transcriptome analysis of the *Arabidopsis* genome nearly a decade ago revealed, surprisingly, that approximately 30% of all the annotated genes exhibited significant asRNA expression [65]. This level

of antisense transcript production has been a consistent observation in other whole-genome analyses, including *Drosophila*, human, and rice [66]. More recently, the application of next-generation sequencing methodologies has revealed the presence of pervasive transcription and long non-coding RNAs (lncRNAs, non-protein-coding RNA molecules at least 200nt in length) in many genomes [67–70]. Transcription is clearly not limited to protein-coding genes and can be antisense to genes, divergent from promoters, convergent into genes, and completely intergenic [66]. The extent of lncRNAs in a genome seems to correlate with an increase in an organism's complexity. For example, the protein-coding part constitutes only 1.5% of the human transcriptome, whereas in prokaryotes it accounts for over 75%. This has led to the suggestion that non-coding parts of a genome serve a gene regulatory role, which becomes increasingly important in higher organisms [71]. lncRNA can play important roles in gene regulation acting as molecular signals, decoys, guides, and scaffolds [69]. It is unclear at present whether cis-localized antisense transcripts, a subset of the more general long non-coding transcripts, play a functionally distinct role [72]. asRNAs have been proposed to function through transcriptional interference (a direct negative impact of one transcriptional activity on a second transcriptional activity *in cis*, i.e. at the same locus) [73], mediate chromatin modifications [74], or be involved in RNA interference mechanisms [75]. RNA interference is a mechanism where occurrence of double-stranded RNA leads to the action of Dicer a nuclease protein complex that cleaves the double-stranded RNA, which is then processed further to small RNAs [1]. In yeast, they have been implicated in stress responsiveness and the regulation of inducible genes [76, 77]. In addition, as a result of their having lower baseline levels, genes expressing asRNA generally exhibit a wider expression range than those genes which do not express asRNA [78]. These characteristics and their abundance in a range of organisms suggest they could play a role in plasticity of gene expression. In plants, this could be of profound importance for responses to environmental cues and adaptation. However, until we determine more mechanistic details on a range of examples from different organisms that will enable us to infer generic

mechanisms, their widespread role in genome regulation will remain an interesting possibility only.

### 1.1.3 *Antisense-mediated regulation in plants*

The presence of extensive antisense transcription in several whole-genome transcriptome analyses was initially viewed with some scepticism and concern that the methodologies were subject to technical artefacts resulting from spurious second strand complementary DNA synthesis in the reverse transcription reaction. This issue has been addressed directly in several studies [79, 80] and extensive antisense transcription is still observed (although less than the early studies). The need for plants to constantly adapt to changing conditions makes a role for asRNA in gene regulation an attractive hypothesis [81], and it is clear from a genome-wide study that antisense transcripts in *Arabidopsis* show developmental, clock-related and stress-related changes in expression [72, 82]. The presence of asRNA has also been detected in rice, wheat, and legumes [83–85], but apart from an analysis of the role of the nonsense-mediated decay pathway on their stability, relatively little mechanistic understanding is available [86]. Given the extensive understanding of small RNA (sRNA) pathways in plants, initial studies addressed whether natural antisense transcripts (NATs) transcribed from either the same locus or a convergently transcribed gene (so-called cis-NATs) triggered a double-stranded RNA mechanism involving sRNA production. However, cis-NATs do not show an increased sRNA signature compared to other genomic regions, so any correlation between the presence of asRNA with low sense expression must involve a different mechanism [87, 88].

To date, the best understood example of the role of antisense transcripts in plants is the regulation of the *Arabidopsis* floral repressor gene *FLC*. Here, antisense expression independently intersects with two pathways that repress *FLC* expression (Figure 1.2). Below we detail our current understanding of this regulation.

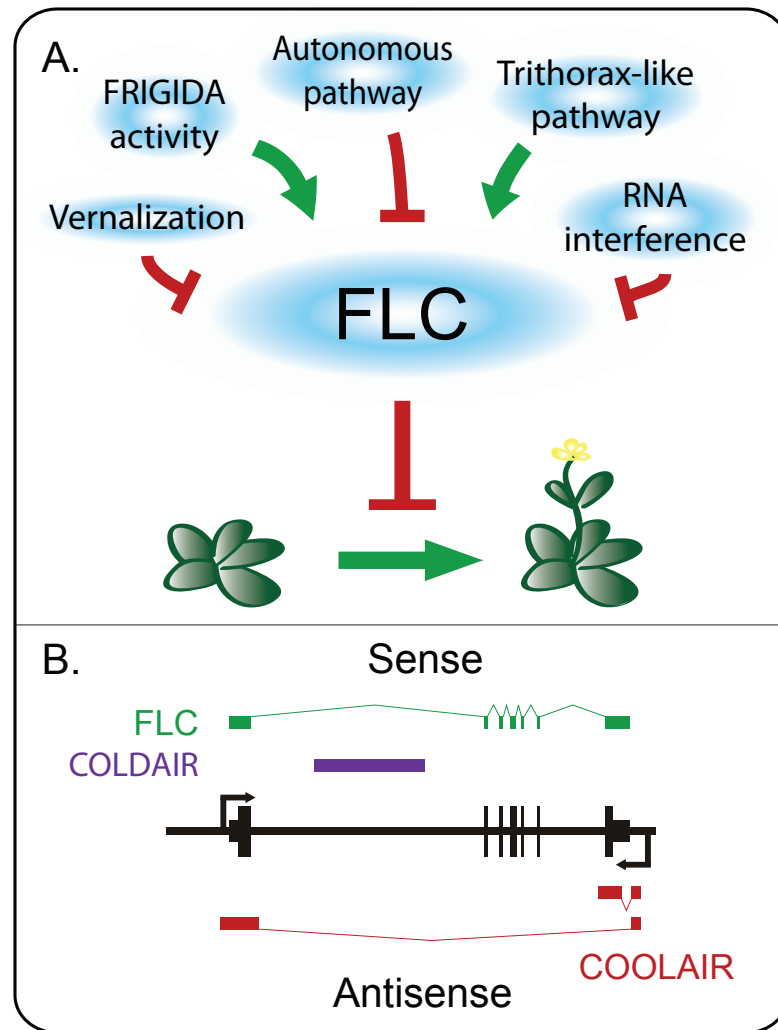


Figure 1.2: The *FLC* gene. (A) *FLC* controls the transition from vegetative to reproductive development in *Arabidopsis* [89–92]. *FLC* expression is regulated by many pathways [29]: vernalization, cold-induced epigenetic silencing that occurs during winter; *FRIGIDA*, a coiled coil protein that upregulates *FLC* expression; the autonomous pathway, which is composed of many independent repressive activities; the *Arabidopsis* Trithorax-like pathway, which stimulates *FLC* gene expression; and RNA interference which suppresses *FLC* in different *Arabidopsis* accessions [93, 94]. (B) The *FLC* locus ( $\sim 6\text{kb}$  in length) expresses multiple types of transcripts (coloured blocks indicate exons, broken lines introns). *FLC* mRNA encodes a MADS box transcription factor. *COOLAIR* is a non-coding transcript that fully encompasses *FLC* in the antisense direction. It is alternatively polyadenylated with a proximal poly(A) site in sense intron6 and a distal poly(A) site in the sense promoter region and is differentially expressed in warm and cold conditions. *COLDIAIR*, expressed under cold conditions from within intron1 of *FLC* in the sense direction, is a capped but non-polyadenylated lncRNA.

#### 1.1.4 Introduction to *FLC* regulation

*FLC* encodes a MADS box-type transcriptional repressor that prevents the activation of a suite of genes required for the floral transition [89–92]. The *FLC* antisense transcription unit fully encompasses the *FLC* gene (~ 6kb in size), initiating from just downstream of the major sense poly(A) site and terminating upstream of the sense initiation site (Figure 1.2). The RNA starts from several positions within a chromatin region marked by the chromatin modification histone H3 lysine 9 dimethylation (H3K9me2), normally associated with heterochromatin in *Arabidopsis* and other organisms. This region also shows homology to sRNAs: a Dicer-dependent (24nt) and a Dicer-independent 30-mer [93]. Analysis of the origin of these sRNAs led to the identification of an antisense transcript that, due to its upregulation by cold, was called *COOLAIR*, as a parallel to the lncRNA *HOTAIR* in *Drosophila* [95, 96]. *COOLAIR* is now used to describe the *FLC* group of antisense long non-coding transcripts generally, both in warm- and cold-treated plants. These antisense transcripts terminate at either proximal sites corresponding to *FLC* sense intron 6, or distal sites within the *FLC* promoter region [97, 98]. This led to the question of whether its initiation depends on the termination of the sense transcript. A fusion of the *COOLAIR* promoter to a luciferase coding region demonstrated that antisense transcription could be initiated and regulated independently of the sense transcript [95]. A similar independence of sense and antisense transcription has been found for yeast genes [99]. *COOLAIR* transcription is generally positively correlated with sense transcription in a range of flowering mutants that either upregulate (late-flowering mutants) or downregulate (early-flowering mutants) *FLC* expression. However, this symmetry is broken during cold treatment: after two weeks of cold treatment, antisense levels accumulate 10 fold, whereas *FLC* transcription is downregulated.

The *FLC* genetic regulatory network is complex, but a schematic illustration of the multiple activating and repressing pathways is shown in Figure 1.2. The activating pathways involve conserved chromatin regulators including homologs of Paf1C [100], RAD6-BRE1 [101–103], and COMPASS [104–106], that carries *Arabidopsis* Trithorax

homologs that trimethylate H3K4 [107]. SWR1, a conserved chromatin remodeler implicated in histone H2A.Z deposition, is also required for *FLC* up-regulation [108]. In addition, the transcriptional activator FRIGIDA contributes strongly to natural variation of *FLC* expression, and thus flowering [109, 110]. FRIGIDA is a coiled-coil, lysine-rich protein that interacts with a CAP binding complex subunit *in vivo* and is required for the recruitment of chromatin modification machinery to the *FLC* locus [108, 110, 111]. SDG8, an H3K36me3 methyltransferase, deposits this histone modification at *FLC* [112, 113] and interacts with FRI to upregulate *FLC* expression [111, 112]. These activators are antagonized by a number of repressors identified by genetic analysis of late-flowering mutants. These were grouped into the so called autonomous pathway. However, rather than functioning in a linear genetic pathway, they constitute several parallel mechanisms that all repress *FLC*.

#### 1.1.5 *Autonomous pathway*

The autonomous pathway has for a long time been known to involve RNA regulation since the autonomous pathway gene *FCA* was shown to encode a protein that contains RNA recognition motifs (RRMs) [114]. *FCA* also has a WW protein interaction domain and this was used to identify the interacting protein FY, previously characterised through genetic analysis of a late-flowering mutant. FY is homologous to Pfs2p (yeast) and WDR33 (mammals), components of the well-characterized pre-mRNA 3' end cleavage and polyadenylation stimulating factor (CPSF) [115–117]. *FCA* and FY were shown to be functionally important in RNA 3' processing as evidenced by their role in autoregulation of *FCA* poly(A) site choice [118] and their genome-wide effect on polyadenylation [119]. In the autoregulation of *FCA*, the *FCA* protein binds to a promoter proximal poly(A) site to stimulate proximal processing, resulting in a truncated transcript [118]. The *FCA* protein is generated from translation of a distal transcript isoform [118]. Through this negative feedback loop, *FCA* tightly controls its own expression levels and thereby *FLC* regulation [118]. Loss of *FCA* also leads to genome wide transcriptional readthrough beyond canonical poly(A) sites into intergenic regions [119]. This results in alternative splicing and polyadenyla-

tion of many transcripts [119]. However, mutations in FCA and FY do not influence *FLC* sense transcript 3' processing.

In order to understand how FCA activity results in a ~ 25 fold reduction in *FLC* expression, suppressor mutagenesis was undertaken to identify other required factors [120]. This identified FLD, a homolog of the human lysine specific demethylase 1 (LSD1, also known as KDM1A) [120, 121] and two canonical 3' processing factors, CstF64 and CstF77 [98]. LSD1 is essential for mammalian embryonic development and is involved in diseases such as leukemia [122, 123]. It has specificity for H3K4me1/2 demethylation and acts as a gene repressor [121]. LSD1 has also specificity for demethylation of H3K9me1/2, in which case it is associated with gene activation [122, 123]. Loss of FLD leads to a genome-wide increase in H3K4me2 levels and also increased H3K4me2 in the *FLC* gene body, confirming its role as a H3K4me2 demethylase [120]. Surprisingly, mutation of the conserved components CstF64 and CstF77, which are normally required for gene expression, actually increased *FLC* levels [98]. Analysis of 3' processing of *COOLAIR* provided insight into the mechanism; these mutations were affecting the processing of the antisense transcript, which in turn increased *FLC* expression.

*COOLAIR* is alternatively polyadenylated (Figure 1.3), as are many transcripts in the *Arabidopsis* genome [124], with both a proximal poly(A) site (within sense intron 6) and a distal poly(A) site (overlapping the sense promoter). FCA and FY function to promote the use of the proximal site, and FCA associates with *FLC* chromatin in this region [120]. Potentially, FCA could interact with *COOLAIR* through its RRM, bringing that transcript to the RNA 3' processing machinery through the interaction with FY [125, 126], thereby stimulating use of the proximal poly(A) site (Figure 1.3). This view is supported by the fact that PRP8 and CDKC2, also identified in FCA suppressor screens, both require intact *COOLAIR* to repress *FLC*. PRP8 promotes splicing of a small intron in *COOLAIR* that enables use of the proximal polyadenylation site [127]. CDKC2, a component of the P-TEFb elongation factor, is required for efficient *COOLAIR* transcriptional elongation [128]. We cannot exclude the

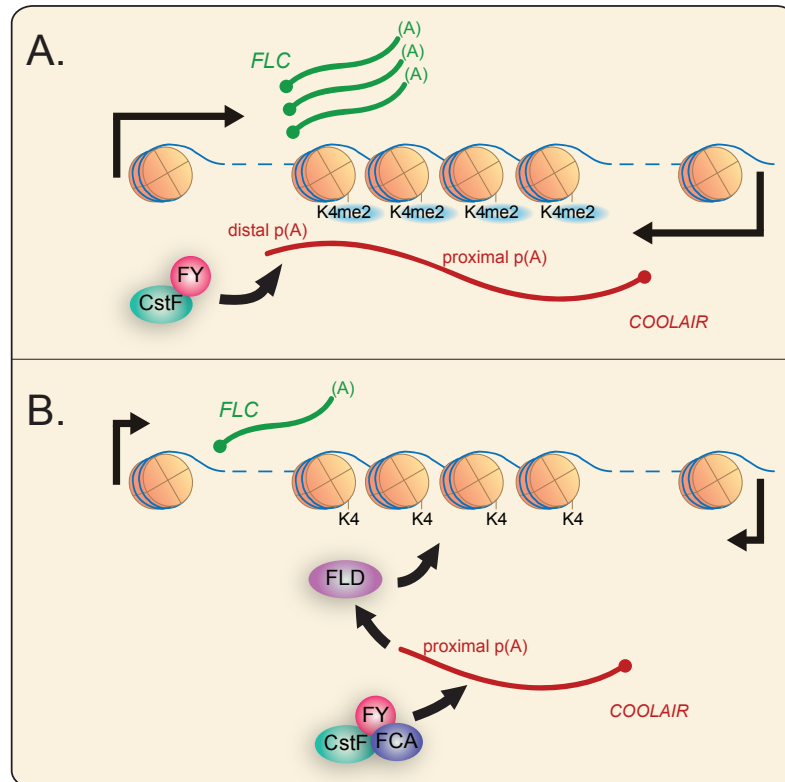


Figure 1.3: The autonomous pathway represses *FLC* expression through asRNA-mediated chromatin modifications. (A) In the absence of endogenous FCA, *COOLAIR* transcripts (red) are polyadenylated under the influence of FY, CstF64, and CstF77 at a distal poly(A) site. This is associated with high H3K4me2 (me2 in blue) levels in the body of the *FLC* gene and high levels of functional *FLC* mRNA (green) and *COOLAIR* expression. (B) FCA promotes alternative polyadenylation of *COOLAIR* through targeting FY and CstF activity to a proximal poly(A) site. This event is likely to trigger FLD-dependent demethylation in the body of *FLC* leading to a transcriptionally repressed state and low levels of functional *FLC* mRNA [120].



possibility that DNA also plays a role in FCA recruitment.

Another feature of endogenous *COOLAIR* is the existence of an R-loop that forms at the *COOLAIR* 5' region [129]. Interestingly, some RRM-domain proteins bind to both RNA and single-stranded DNA, a feature associated with R-loop formation [130]. Thus, FCA may be recruited through a combination of sequence specificity and nucleic acid structure. FPA, another RRM protein that acts independently of FY [131], also promotes usage of the proximal site [97]. The choice of the proximal poly(A) site could then trigger FLD demethylation to reduce the H3K4me2 levels in the body of *FLC* leading to a repressed chromatin state and reduced sense and antisense expression [98, 120]. How use of the proximal poly(A) site stimulates FLD activity remains to be elucidated. A relative increase in use of the distal poly(A) site in *fca*, *fpa*, or *fy* mutants (Figure 1.3) has led to the suggestion that increased antisense transcription through the sense promoter could stimulate sense transcription [97, 132]. Opposing proximal poly(A) and distal poly(A) site choice could thus function antagonistically as an *FLC* repressor and activator, respectively. In this way regulated 3' processing of the *FLC* antisense transcript, *COOLAIR*, might modulate sense *FLC* expression. More generally, how the overall chromatin state mechanistically affects sense *FLC* expression is unknown. The conserved nature of several involved proteins suggests this could be a mechanism generally relevant for the many loci with antisense transcription.

#### 1.1.6 Vernalization

The autonomous pathway functions in parallel with vernalization, a second pathway repressing *FLC* expression [133]. Vernalization is a process whereby flowering is accelerated by prolonged cold, and it ensures that plants align their flowering with spring [29]. Prolonged cold represses *FLC* transcriptionally and induces epigenetic silencing that is mediated by a conserved Polycomb (PcG) mechanism [134]. *FLC* expression decreases in the cold in a quantitative manner, scaling with the length of cold that the plants experience, with *FLC* remaining epigenetically silenced during subsequent development after return to the warm [135]. Mutational analysis revealed that this

process requires a modified PcG silencing complex including proteins containing a plant homeodomain (PHD). A core Polycomb repressive complex 2 (PRC2) associates with *FLC* chromatin independently of temperature [136]. After several weeks of cold, levels of histone H3 lysine 27 trimethylation (H3K27me3), which is associated with PcG silencing, accumulate at an intragenic site covering (sense) exon 1 and the 5' end of intron1, referred to as the nucleation region (Figure 1.4). This correlates with accumulation of the modified PHD-PRC2 complex and its association with the nucleation site. H3K27me3 accumulates quantitatively at the nucleation region with increasing weeks in the cold [134, 137]. Upon return to warm temperatures, the PHD-PRC2 complex spreads across the whole *FLC* locus raising H3K27me3 to high levels, which are required to maintain the epigenetic stability of repression through subsequent cell divisions and environmental noise. This switch to the fully epigenetically silenced state is digital: it either occurs and the gene is fully silenced; or it does not occur and the gene is re-activated upon return to warm temperatures [138]. The probability of this cell-autonomous switch increases with length of cold so the quantitative accumulation of epigenetic silencing by cold represents an increasing fraction of cells that have switched to the epigenetically silenced state [138, 139]. The question then arises as to how *COOLAIR* function in this mechanism?

Using a custom microarray with single nucleotide resolution of both strands of the *FLC* locus, changes in *FLC* transcripts at several phases of the vernalization process were identified [95]. The most striking difference was the accumulation of *FLC* antisense transcripts after plants had experienced cold, peaking after 14 days cold treatment. Unspliced and proximally polyadenylated *COOLAIR* increased most significantly (~ 10 fold), but the distally polyadenylated form also increased slightly. The increase in antisense transcripts coincides with suppression of unspliced sense *FLC* levels, but not functional *FLC* mRNA levels, which take several more weeks of cold to decrease significantly. This led to the idea that *COOLAIR* may be facilitating the decrease in sense *FLC* transcription during the first few weeks in the cold (Figure 1.4). Supporting this hypothesis, a transgene that con-

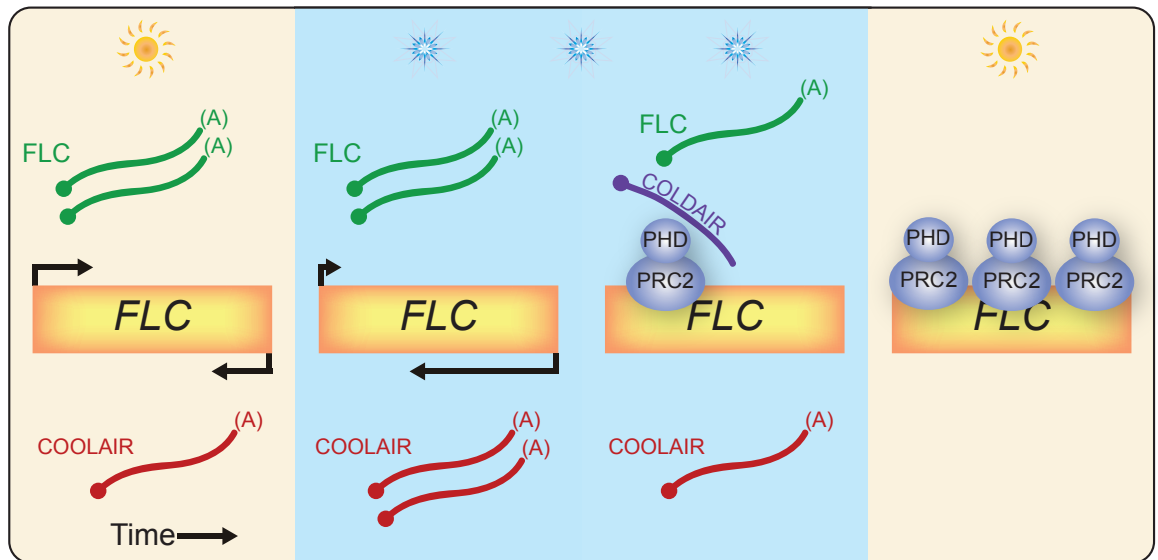


Figure 1.4: Vernalization silences *FLC* epigenetically. Vernalization is the process whereby flowering is accelerated in response to the prolonged cold of winter. Time runs from left to right. Before exposure to cold, *FLC* (green) is expressed in higher amounts than *COOLAIR* (red). During the first two weeks of cold, *COOLAIR* is upregulated while unspliced *FLC* RNA, commonly thought to reflect sense transcription, decreases [95]. However, functional *FLC* mRNA levels are still similar to non-vernalized conditions, possibly due to differences in lifetimes between nascent and processed RNAs. During prolonged cold, *FLC* is progressively epigenetically silenced via a Polycomb-dependent mechanism correlating with *COLDIAIR* expression [134, 138, 140]. Functional PHD-PRC2 complexes associate with the nucleation region slightly downstream of the sense promoter, leading to an up-regulation in the repressive histone mark H3K27me3 at this site. After returning to warm conditions, the PHD-PRC2 complex associates across the whole locus, leading to high levels of H3K27me3 across the gene. High H3K27me3 levels are required for the epigenetic stability of *FLC* repression through the rest of development.

tains the *COOLAIR* promoter fused downstream of a constitutively expressed green fluorescent protein (GFP) showed cold-induced antisense transcript production and cold-induced down-regulation of GFP expression [95]. A higher stability of the spliced *FLC* mRNA compared to nascent transcripts, might explain the observation that *FLC* mRNA levels remain unaffected during a two-week period of cold. When vernalized for longer, *COOLAIR* expression and *FLC* mRNA decrease over time and remain stably silenced after return to the warm [95, 140]. In plants where T-DNA insertions attenuate production of *COOLAIR*, PcG silencing was not disrupted after four weeks of cold treatment, showing that antisense transcripts are not required for vernalization [141]. However, this does not preclude a possible function for *COOLAIR* during the first two weeks of vernalization preceding epigenetic silencing [95]. Indeed, *COOLAIR* accelerates transcriptional shutdown of *FLC* during cold exposure. Removal of *COOLAIR* disrupted the synchronized replacement of H3K36 methylation with H3K27me3 at the intragenic *FLC* nucleation site during the cold [142].

A sense lncRNA (~ 1kb), termed *COLDAIR*, has also been detected within *FLC* intron1 during the cold (Figure 1.2). It contains a 5' cap but is not polyadenylated and is induced by cold, but later than *COOLAIR* reaching maximum levels after 3 weeks of cold. *COLDAIR* was found to be associated with the core PRC2 components, and its knock-down led to attenuation of epigenetic silencing, suggesting it plays an important role in triggering PcG silencing [140]. In summary, there appears to be a complex interplay between antisense transcripts, other lncRNAs, and chromatin-modifying complexes in the various steps of the vernalization process (Figure 1.4). It will take integration of genetic, biochemical, and modelling approaches to unravel the mechanistically redundant functions contributing to this mechanism.

Our overall understanding of antisense-mediated gene regulation is in some cases further advanced in yeast and mammals. Below we will discuss some well-studied examples below, which provide important concepts for the broader roles of asRNA in chromatin regu-

lation. To gain further insight into this connection, we will suggest some possible mechanistic links between *FLC* and these systems.

#### 1.1.7 *Genetic toggle mechanism*

A universal issue in asRNA-chromatin regulation is to understand whether the RNA molecule itself or the act of transcription is important for the regulation [143]. The dissection of regulation of the yeast *FLO11* gene suggests that for some examples it is the latter. *FLO11* encodes a glycoprotein important for adhesion, and it is regulated by a pair of antagonistic lncRNAs resulting in a variegated gene expression pattern that is likely to give considerable selective advantage to cells as the environment changes [63, 144]. A sense lncRNA (*ICR1*) is transcribed upstream of the *FLO11* promoter. Antisense to this is another lncRNA, *PWR1* (Figure 1.5). Competitive binding of the activator (Flo8) or repressor (Sfl1) to the *FLO11* promoter determines which of the two lncRNAs is transcribed. Binding of Flo8 stimulates *PWR1* transcription and inhibits *ICR1* expression, establishing a basal expression state of *FLO11*, which can be upregulated by conventional transactivating factors. *ICR1* transcription is thought to clear the *FLO11* promoter, enabling binding of Sfl1, which then recruits an HDAC to fully silence the gene. The overall effect is variegated transcription in individual cells, monitored in a recent study using single-molecule RNA fluorescence in situ hybridization (FISH) [63, 145], with three expression states: silenced, basal, or active expression. The promoter-localized lncRNAs, one of course antisense to the other, thus provides a genetic toggle that contributes to the observed variegated expression of *FLO11*. This concept is appealing in the *FLC* context: instead of a toggle mechanism determining which promoter lncRNA is expressed as in the case of *FLO11*, the choice of either the proximal or distal antisense poly(A) site could be a toggle mechanism for *FLC*. This switch may then be the signal for recruitment of chromatin machinery necessary for low or high *FLC* expression states.

#### 1.1.8 *RNA turnover linked to histone modifications*

Another classic example of a functionally important antisense transcript in yeast is that corresponding to the *PHO84* gene, which en-

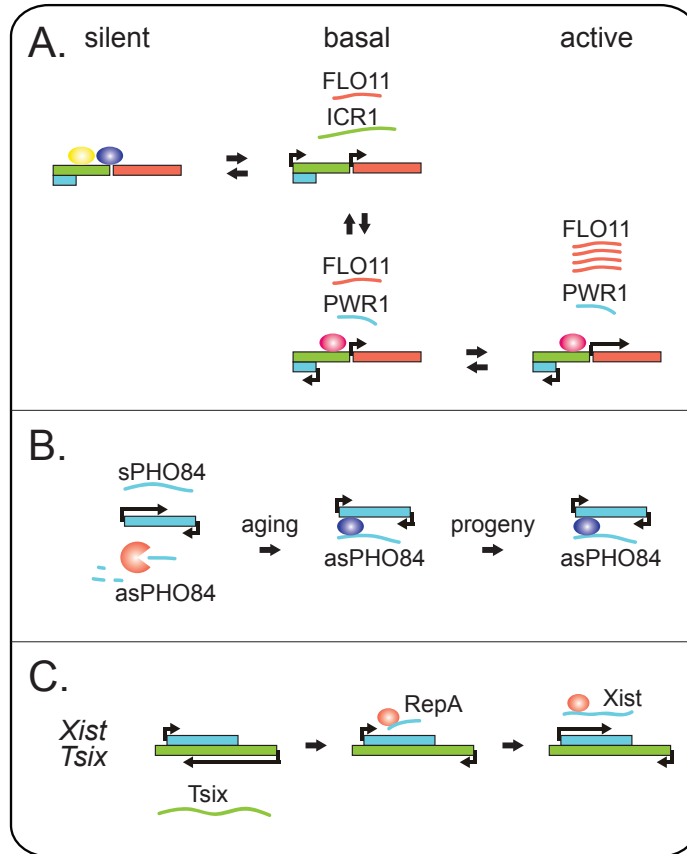


Figure 1.5: Antisense RNA can regulate gene expression in yeast and mammals. (A) *FLO11* (red gene) expression in yeast can be in three states: silent, basal, or active. The silent state is associated with binding of repressor Sfl1 (yellow protein), which recruits an HDAC (purple protein) to the promoter, leading to transcriptional silencing of *FLO11*, the sense ncRNA *ICR1* (green gene), and asRNA *PWR1* (antisense direction, blue gene). In the basal state, expression of the ncRNAs toggles between *ICR1* and *PWR1*, a process associated with basal low expression of *FLO11*. *ICR1* overlaps with the *FLO11* promoter, and its transcription is thought to clear the promoter of binding proteins, allowing competitive binding between Sfl1 and the activator Flo8 (pink protein). Conversely, *PWR1* expression interferes with *ICR1* and is associated with *FLO11* activation because it can synergize with Flo8 binding and recruit further transcriptional machinery. This results in an active state with high *FLO11* levels [63, 145]. (B) Another yeast gene, *PHO84* (blue gene), is epigenetically silenced in the process of chronological aging. Sense expression (*sPHO84*) is initially high. As a result of reduced nuclear exosome activity over time, higher anti-sense *PHO84* (*asPHO84*) expression induces recruitment of HDACs (purple protein) to the sense promoter. This epigenetically silenced state is inherited by the progeny [74]. (C) Before the initiation of female X-chromosome inactivation (XCI), *Xist* (blue gene) is silenced because *Tsix* (green gene) expression induces DNA methylation at the *Xist* promoter. At the initiation of XCI on the chromosome that is to be inactivated, *Tsix* is downregulated, and it is proposed that *RepA*, a lncRNA located within *Xist* in the sense direction, targets PRC2 complexes (red protein) to the locus. At the inactivated X chromosome, *Xist* is expressed and associates with PRC2 complexes and spreads in a cis-limited fashion across the inactivated X chromosome [146, 147].

codes a phosphate transporter. *PHO84* asRNA originates from a convergent promoter at the 3' of the gene (Figure 1.5). Chronological ageing of yeast cells, a process influenced by cold, was found to induce this asRNA, which resulted in silencing of *PHO84* sense transcription in an epigenetically stable manner [74]. This arose from inactivation of the nuclear exosome over time, which enabled the asRNA transcripts to accumulate and mediate HDAC recruitment to the *PHO84* promoter leading to epigenetic silencing. These data suggest it is the RNA itself that is important for this mechanism. Introduction of an ectopic *PHO84* gene copy led to silencing of both the endogenous and ectopic gene mediated by its asRNA. However, this was via a somewhat different mechanism not requiring HDACs but involving the H3K4 methyltransferase Set1 that stimulates antisense transcription [148]. In yeast, chronological ageing reduces exosome activity, stabilizing antisense transcripts.

The *PHO84* system raises the question: could environmental factors alter exosome function thereby triggering chromatin modifications and gene expression changes? It will be interesting to determine if cold during the first few weeks of vernalization functions through such a mechanism to upregulate *COOLAIR*.

Different RNA turnover mechanisms appear to delineate functional ncRNA in yeast [149]. One class of over 1000 ncRNAs antisense to yeast ORFs are degraded by the cytoplasmic 5'-3' exoribonuclease Xrn1 [76]. These Xrn1-sensitive unstable transcripts (XUTs) are polyadenylated and transcribed by Pol II, for example at the *TIR1* locus and retrotransposon *TY1* [150]. In *xrn1* mutants, *TIR1axut* and *RTL* (the antisense transcripts of *TIR1* and *TY1*, respectively) are stabilized and this correlates with decreased sense gene expression. *RTL* also mediates chromatin silencing *in trans*, i.e. not at the same locus where *RTL* was transcribed, partially through histone deacetylation and histone methylation [150], with H3K4me and H3K4me2 being important for XUT-mediated silencing [76]. H3K4me3 antagonizes XUT activity and this may result in the antisense transcripts mediating gene silencing in a chromatin dependent manner: in the presence of elevated H3K4me3 levels at the sense promoter, a hallmark of ac-

tive chromatin, stabilization of XUTs seems to have less of an effect on sense gene expression [76]. This reinforces the suggestion that asRNA-mediated repression predominates when sense expression is already low, leading to an increased expression range due to lower baseline levels in the repressed state [78].

#### 1.1.9 *Antisense RNA in Polycomb silencing*

There has also been a large amount of analysis on the role of specific sense/antisense lncRNAs in chromatin regulation in mammals, the most well-known pair being *Xist* and *Tsix*. These lncRNAs play a central role in the choice of which X chromosome is inactivated in female mammals [146, 147]. *Xist* encodes an 18kb nuclear localized RNA essential for X-chromosome inactivation (XCI). Before the initiation of XCI, *Xist* is antagonized by biallelic transcription of *Tsix*, its antisense counterpart. This silencing is not due to occlusion of the transcription machinery at the *Xist* promoter by *Tsix* transcription (as suggested in the genetic toggle mechanism described above) but rather by *Tsix* RNA associating with DNA methylation machinery to stably silence the *Xist* promoter. It is proposed that during X-chromosome pairing, down-regulation of *Tsix* on the future inactive X chromosome allows *RepA*, a sense lncRNA located within an *Xist* exon, to target PRC2 to *Xist*, which remains activated during XCI. *Xist*-PRC2 complexes tethered to the chromatin then spread in a cis-limited fashion to cover the whole X chromosome, resulting in its inactivation.

Similar principles of transcription from both strands, chromosome coating, and association with chromatin machinery has also been demonstrated for lncRNAs involved in genomic imprinting in mammals [143, 151, 152]. These RNAs silence a set of genes in an allele specific manner such that these genes inherited through the maternal and paternal gametes are differentially expressed. A well-characterized example is *Kcnq1ot1*, which encodes a 91kb transcript antisense to the *Kcnq1* gene cluster. This RNA coats a ~ 1Mb region of the paternal chromosome, resulting in the epigenetic silencing of 8-10 genes in this *Kcnq1* domain that remain expressed from the maternal alleles.



From the discussion above describing *Xist*, *Tsix*, *RepA*, and *Kcnq1ot1* as well as the involvement of *COOLAIR* and *COLD AIR* in regulating *FLC* during vernalization, it is clear that sense and antisense lncRNAs are involved in PcG-mediated silencing of genes in both plants and mammals. Although the details of each regulatory system appear different at this stage, further conceptual parallels are likely to emerge from further investigation.

We have described above the key features of eukaryotic transcriptional regulation, which occurs in the context of a chromatin environment. This is relevant to understand the complexity of *FLC* regulation, which, as described above, also involves a non-coding antisense transcript *COOLAIR*. Therefore we have included a description of our understanding of antisense lncRNAs in plant gene regulation, with comparison to well-characterized examples in yeast and mammals. In Chapter 2, we will describe our primary research findings on the quantitative regulation of *FLC* by the autonomous pathway and *COOLAIR*. In Chapter 3 we then proceed with investigating the cellular variability in *FLC* levels and *FRI*-mediated *FLC* regulation. The quantitative variation in expression that antisense regulation confers could be very important in many natural contexts. With respect to flowering time control, subtle changes in timing have large implications for seed yield and thus reproductive success. We might see many environmental inputs influencing the antisense-mediated regulation of *FLC* in specific genotypes and also find extensive variation in this mechanism in natural populations adapted to very different climates.

## 1.2 Regular positioning of low copy number plasmids

### 1.2.1 DNA segregation in prokaryotes

In many living organisms stable DNA inheritance is crucial to proliferation [1, 153]. Intricate mechanisms have evolved to ensure that the genome is accurately moved and positioned from parent to daughter cells. In prokaryotes genetic material comes in several forms [1,

153]. Most common are chromosomal DNA and plasmids. Plasmids are often small ( $1 - 10^3$  kbp) compared to chromosomes (4.6Mbp in *Escherichia coli*, *E. coli*) and they replicate in a cell-cycle independent manner [1, 154]. Some plasmids occur in high copy number and generally partition randomly at cell division [154]. Other plasmids only occur at low copy numbers ( $1 - 10$ ) and exhibit active segregation mechanisms [153, 155–159]. Therefore the latter represent good model systems to study segregation of genetic material. Furthermore many plasmids carry virulence factors or antibiotic resistance genes [156, 160], further highlighting the relevance to understand the mechanisms underlying their DNA inheritance.

In eukaryotes chromosome separation during mitosis occurs under the influence of the mitotic spindle, of which microtubules are an important constituent [1]. However it remains unclear to what extent prokaryotes possess an intracellular cytoskeleton and if it is required for DNA segregation [155]. In the last decade, much progress has been made to elucidate the mechanisms underlying DNA segregation in prokaryotes. These partitioning systems can be broadly classified into three classes [155]. The most well-characterised type is that encoded by a *parMRC* operon, encoding for the actin homolog ParM and DNA binding protein ParR, that binds to a *cis* DNA element termed *parC* [155, 161, 162]. In this case, actin-like filaments, reminiscent of the mitotic spindle, push plasmids apart through polymerization [163, 164]. After reaching the pole, these filaments completely depolymerize [163]. Another class of DNA segregation mechanisms is based on the more enigmatic tubulin homolog TubZ [165]. Like tubulins, that are core subunits of microtubules, TubZ performs treadmilling activities of polymer growth and depolymerization that seem important for plasmid segregation [165]. The last and most common class of prokaryotic DNA partitioning mechanisms is encoded by a *parABC* locus [155, 159, 166–168]. In several bacteria, for instance in *Vibrio cholerae* (*V. cholerae*) and *Caulobacter crescentus* (*C. crescentus*), this machinery segregates chromosomes prior to cell division [167–169]. The *parABC* locus is also present in *E. coli* plasmids such as pB171 and P1 [159, 166] and antibiotic-resistance-

carrying plasmids [160]. In the next section we detail our current understanding of this ParA family of partitioning mechanisms.

### 1.2.2 *The ParA family of plasmid partitioning mechanisms*

The *parABC* operon, present in *V. cholerae* and *C. crescentus* chromosomal DNA and in plasmids in a variety of different bacteria, encodes for two proteins [159, 166]: ParA, a P-loop ATPase that dimerizes in its ATP-bound form (ParA-ATP for short), and the DNA-binding protein ParB that binds site-specifically to the neighboring *parC* region [170, 171]. ParB also homo-dimerizes [171]. How many ParB (dimers) bind to the *parC* region depends on the specific partitioning system, and it is unclear if this number is functionally important for inheritance in every system. For plasmid pB171 it is clear that ParB binding is limited by the 35 binding sites divided over two *parC* region [170], *parC1* upstream of the *ParA* and *ParB* and *parC2* immediately downstream [156]. On the other hand, plasmid P1 ParB protein can spread for several kilobases along flanking DNA and which is thought to have an auxiliary role in plasmid partitioning [172].

ParA-ATP binds DNA non-specifically in its dimeric ATP-bound form *in vitro*, a feature that is not dependent on its ATPase activity [166]. *In vivo* ParA also colocalizes with the nucleoid [159, 167, 173, 174]. Fluorescence microscopy has provided evidence for plasmid pB171 ParA movement over the nucleoid with spatiotemporal oscillations in helix-like structures [156–158]. ParB and *parC* are required for these dynamics [156], with ParB promoting the conversion of ParA-ATP to dimeric ParA-ADP (ParA-ADP for short), causing ParA to unbind from the nucleoid [159, 166]. The time period required for nucleoid-disassociated ParA to regain the ability to bind the nucleoid is sufficiently long *in vitro* to ensure that the relative locations of ParA-ADP unbinding and later ParA-ATP re-binding would be uncorrelated due to cytoplasmic ParA diffusion [166]. In contrast, P1 ParA-ATP does not oscillate and forms a gradual distribution, with occasionally an accumulation of ParA around the plasmid [173]. F Plasmid ParA (SopA) exhibits a ParA focus that

oscillates from pole-to-pole and associates with filamentous ParA structures on the nucleoid [175]. However, *in vitro* SopA does not oscillate and seems to form a graded distribution [176], similar to P1 ParA [177]. Plasmid pB171, F and TP228 ParA-ATP polymerize strongly *in vitro* [158, 178, 179], whilst P1 ParA merely dimerizes [166, 176, 180]. Due to these seemingly contradictory observations, it is currently controversial whether ParA-ATP polymerizes to form long filaments *in vivo*. It seems the degree of polymerization is dependent on the specific *parABC* system [181]. It is currently unclear whether the degree of polymerization is functionally important for plasmid segregation. A similar debate on the degree and relevance of polymerization is also ongoing for chromosomal ParA segregation [167–169] despite evidence that they form filamentous structures *in vivo*.

Interestingly, *parABC* loci generate equally spaced positioning of plasmids along the long axis of the cell (Figure 1.6), thereby ensuring stable plasmid inheritance [157–159, 183]. In cells with only one plasmid, it seems localized in the middle of the cell. In the presence of two plasmids, the localization pattern is different: in this case plasmids reside at around 1/3 and 2/3 along the long cell axis. This generalizes further to higher plasmid copy number. All three components, ParA, ParB and *parC*, are essential for positioning [157–159, 183]. Regular positioning occurs in many bacterial low copy number plasmids, including pB171, P1 and F plasmids [157–159, 175, 183]. However, the underlying mechanism of action is not satisfactorily understood. Several questions are outstanding. Firstly, the means by which plasmids translocate under the influence of ParA, and secondly whether ParA polymerization is important for this movement. Several mechanisms have been proposed to explain ParA-mediated plasmid motion [184]. One hypothesis proposes that ParA-ATP polymerizes on the nucleoid to form long filaments and that plasmid translocation is achieved by ParB-stimulated retraction of the polymers, generating effective plasmid-pulling [159, 168]. Brownian dynamics simulations indicate that ParB-*parC*-mediated disassembling ParA polymer bundles can indeed both tether and pull plasmids simultaneously without the need for plasmid diffusion

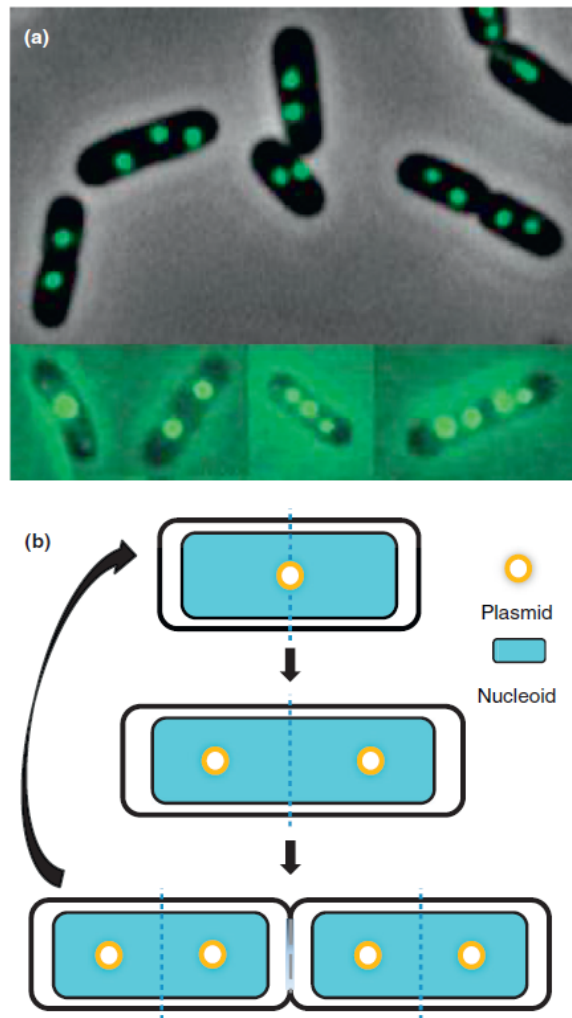


Figure 1.6: Regular distribution of plasmids on the bacterial nucleoid confers genetic stabilisation of plasmids by *parABC* loci. (A) Live cell images showing plasmids (green) are distributed by their respective *parABC* locus at regular intervals over the nucleoid (plasmids P1 in upper panel and pB171 in lower panel) in *E. coli* cells (black). (B) Schematic representation of the continuous adjustment of plasmid (orange circle) positioning over the nucleoid (blue rectangle). As the plasmids replicate while the cell cycle progresses in parallel, the *parABC* locus adjusts inter-plasmid distances and ensures a dynamic, but on time-average, even distribution. It is currently unclear how this equal spacing of plasmids over the nucleoid is generated. This illustration is reprinted from [182] with permission from Elsevier.

[185]. Other proposals are based on ParA-ATP forming a gradient-like distribution on the nucleoid, without a necessity for polymerization [166, 169, 173, 176, 177, 180, 186]. In this so-called diffusion-ratchet mechanism [166, 176, 177, 180], it has remained unclear how ParA actually influences plasmid movement, e.g. through plasmid diffusion or actively directing plasmid motion through a chemophoresis force [176, 186]. Another proposed mechanism is a DNA-relay for *C. crescentus* chromosomal *parC* motion from pole to pole upon DNA replication [169]. A pulling force would then be generated by the elastic dynamics of the nucleoid DNA to which ParA-ATP dimers are bound [169]. However, using super-resolution microscopy *C. crescentus* ParA has also been observed to form narrow filaments, which lead to the proposition of long polymer filaments along which *parC* perform a 'burnt-bridge' biased diffusion [167]. In such a scenario, the *parC* region would diffuse along the ParA filament, whilst depolymerizing it through ParB-mediated ParA ATPase stimulation.

It is currently unclear whether any of the above described mechanisms can explain equal plasmid spacing given the known physiological and biochemical constraints. This is mostly because it has remained unclear what features of the dynamics arising under the influence of the *parABC* operon can deliver the information required to generate equal plasmid spacing. Spacing apparently occurs independently of plasmid copy number and cell length for both pB171 and P1 ParA [159, 183]. In [159] a model based on long ParA polymers pulling plasmids was proposed where a length-dependent plasmid drop off rate ensured equidistant positioning. However, it remained unclear what underlying mechanism might cause such a length-dependent off rate. In the diffusion-ratchet mechanism that did not require polymerization, it was suggested that the required information for plasmid spacing would lie in a ParA gradient formed over the nucleoid [166]. Lastly, the relevance of ParA oscillations, whose presence seems to differ between the pB171 and P1 systems [158, 159, 173], to regular plasmid positioning remains to be elucidated.

### 1.2.3 *Regular positioning of other subcellular machinery*

Regular positioning along the long cell axis is not only present in plasmid segregation systems. It is also observed for low copy number protein complexes, such as carbon-fixing carboxysomes [187]. Carboxysomes are microcompartments in cyanobacteria that sequester from the cytoplasm the enzymes responsible for carbon fixation [187]. Similar to plasmids, these spherical compartments are spatially ordered along the long axis of the cell independent of cell length. As a consequence, cells undergoing division segregate carboxysomes in a non-random fashion. This process is also dependent on a ParA protein as its disruption caused positional disordering and impaired carbon fixation. Although ParA seems to oscillate from pole-to-pole, it remains unclear whether this ParA binds to the nucleoid.

Chemotactic clusters in *Rhodobacter sphaeroides* (*R. sphaeroides*) are also spatially positioned in an ordered fashion [174, 188]. These clusters represent a discrete region within the bacterial cytoplasm, similar to carboxysomes, where several of *R. sphaeroides* proteins responsible for chemotaxis reside [188]. Most cells contain either one cluster localized at mid-cell or two clusters at the one-fourth and three-fourths positions of cell [188]. This positioning is dependent on an orphan ParA homolog termed PpfA encoded in the genome. Interestingly, the role of ParB seems fulfilled by TlpT, one of the proteins that also localize in the cluster and is essential for chemotaxis [188]. PpfA regulates the position of these clusters and its number per cell [188]. PpfA has similar DNA binding properties as plasmid ParA proteins and is vital for both cluster positioning and chemotaxis [174].

### 1.2.4 *Cellular architecture of prokaryotes*

Given the DNA binding properties of ParA and its colocalization with the nucleoid *in vivo* [159, 166, 173], it has been proposed that the nucleoid acts as a matrix for the ParA and plasmid dynamics. Interestingly, it has been reported that the circular *E. coli* chromosome adopts a helical shape [189, 190]. Not inconsistent with these findings, another study found that it is highly organized in a linear filament along the long cell axis [191]. Visualization of living nucleoids

suggests that the nucleoid is ellipsoidal [189]. Likewise, the *C. crescentus* is also ellipsoidal, with twisted arms and the chromosomal *parC* site (*parS*) located at one pole [192]. In *E. coli* the nucleoid is radially confined by the cell cylinder and longitudinal density waves flux back and forth along the nucleoid arising from accumulation and relief of intracellular stress [189]. During replication in *Bacillus subtilis*, newly synthesized DNA is translocated via a helical structure from mid cell to the pole [193]. In *E. coli*, the replication origin region (OriC) is generally located around mid-cell, whilst the terminal region (Ter) is more associated with the cell membrane [194]. In *E. coli* MatP is a nucleoid associated protein that helps to compactify and organize the Ter region into a macrodomain, which is enriched for MatP-specific DNA sequence motifs termed *matS* [195]. Loss of MatP results in chromosomal disorganization throughout the cell cycle [195]. The MukBEF complex, that forms foci in *E. coli*, shapes nucleoid organization independently of DNA replication by positioning the OriC and compactifying the nucleoid [196]. Plasmid stability seems unaffected by loss of MukB [197, 198]. It is currently unclear if the nucleoid structure or dynamics are functionally important for plasmid positioning. It could be that the observed helical ParA structures [157] are caused by such a helical nucleoid confirmation. It is clear that, despite the absence of histones [1], many proteins function to generate a highly organized nucleoid [194]

The *E. coli* cytoplasm is also believed to possess glass-like properties, with larger cellular objects being relatively less mobile than their smaller counterparts [199]. This affects plasmids as well as chromosomal DNA loci and other large cellular objects [199]. Metabolic activity is known to enhance cytoplasmic fluidity [199], consistent with the finding that metabolic activity affects plasmid mobility [200]. It is an open question how the *parABC* systems affects plasmid mobility in light of the physical properties of the bacterial cytoplasm. Investigations of the exact involvement of the nucleoid and cytoplasm in intracellular cargo positioning are therefore needed. Intracellular organization more generally is likely to affect many important processes, such as gene regulation and spatial positioning, both in prokaryotes and eukaryotes.



Above, we have described the key known features of regular positioning of low copy number plasmids in bacteria. In Chapter 4, we detail how regular equal plasmid spacing can be achieved, something that has so far remained unclear. We achieve this quantitative understanding through a combination of mathematical modelling and experimenting.

### 1.3 Concluding remarks

In this chapter we have introduced the biological topics of our investigations: *FLC* regulation and regular plasmid positioning by *parABC*. Clearly, we already possess a vast qualitative understanding of many aspects of these systems. However, the aim of my PhD research has been to advance our *quantitative* understanding of them. The two topics are similar in that a qualitative understanding has not elucidated all features and as such many questions remain. As we show in the next three primary results chapters, quantitative modelling can assist in answering many of these questions and help to connect our knowledge of biology with the laws of physics that are relevant to life.



# COORDINATION OF *FLC* TRANSCRIPTION INITIATION AND ELONGATION

---

## 2.1 Introduction

In this chapter, we investigate transcriptional repression of *FLC* by the autonomous pathway. We focus on understanding the roles of FCA, an RNA-binding protein stimulating alternative polyadenylation of *COOLAIR* [98], and FLD, a histone H3 lysine 4 dimethylation (H3K4me2) demethylase [120]. The involvement of FLD has implicated *FLC* chromatin in the *FLC* repression mechanism [98]. It has remained unclear, however, how *FLC* repression through the chromatin environment is mediated and in what way *COOLAIR* is involved in this repression.

We develop several quantitative models for *FLC* repression based on transcription initiation, premature termination or nascent RNA degradation. We use experimental measurements of RNA Polymerase II (Pol II) and total RNA upregulation to distinguish between these hypotheses and parameterize an analytic mathematical model of *FLC* transcription involving coordination of Pol II initiation and elongation. Model predictions are then tested through detailed measurements of intronic total and chromatin-bound RNA. Through our combination of mathematical modelling of experiments, we show that at *FLC* both FCA- and FLD-mediated repression occurs not only through reduced transcriptional initiation but also through a coordinately reduced Pol II elongation rate. Our methodology should be widely applicable for evaluating elongation rate changes in whole organisms where other experimental techniques established for cell culture systems are unfeasible.

Using a stochastic model we also show that the experimental data is quantitatively consistent with a mechanism where slow elongation due to the chromatin state is critical for FCA-mediated proxi-

mal polyadenylation of *COOLAIR*. This would then be a key step in robust *FLC* repression, where *COOLAIR* proximal termination, the local chromatin state and slow Pol II elongation are all required and reinforce each other to achieve stable repression. We propose that coordinated initiation and elongation rate changes could be a general feature of transcriptional regulation connected to the local chromatin state.

## 2.2 Results

### 2.2.1 *FLC expression changes are consistent with regulation of transcription*

In order to investigate *FLC* repression by FCA and FLD, we first quantified how they affect the steady state RNA levels of *FLC*. We found both spliced *FLC* and unspliced *FLC* (measured at introns 2 and 3) increased  $\sim 20 - 25x$  in both the *fca-9* and *fld-4* null mutants compared to the early flowering accession Columbia (Col, wild type) (Figure 2.1). The *fca-9* allele, first described by Chou et. al. [201], contains a basepair deletion, resulting in a reading frame shift [202]. *fld-4* is a Salk T-DNA insertion line [203]. Given the known functions of FCA and FLD described in subsection 1.1.5, there is substantial evidence to suggest that *FLC* is regulated at the transcriptional level. We therefore begin our investigations by focussing on autonomous pathway regulation of *FLC* production. In *fca-9* mutants the proximal antisense upregulation is  $\sim 2x$  and the distal  $\sim 13x$  (Figure 2.1). It has been proposed that this differential upregulation of the two antisense isoforms could be caused by a kinetic coupling mechanism between elongation and splicing [98, 127]. Using an initial system of ordinary differential equations (ODEs, (Equation 2.1)) we quantify that these qualitative notions are consistent with the RNA fold upregulations (Figure 2.1). In subsequent sections, we then proceed with constructing a mechanistic model of transcription by Pol II based on these hypotheses.

We represent sense (antisense) unspliced transcript levels as  $U_s$  ( $U_{as}$ ), spliced sense levels as  $R_s$  and proximal (distal) spliced and polyadenylated antisense as  $R_{as}^p$  ( $R_{as}^d$ ). For both sense and antisense we have respective rates for RNA production ( $p_s, p_{as}$ ), RNA process-

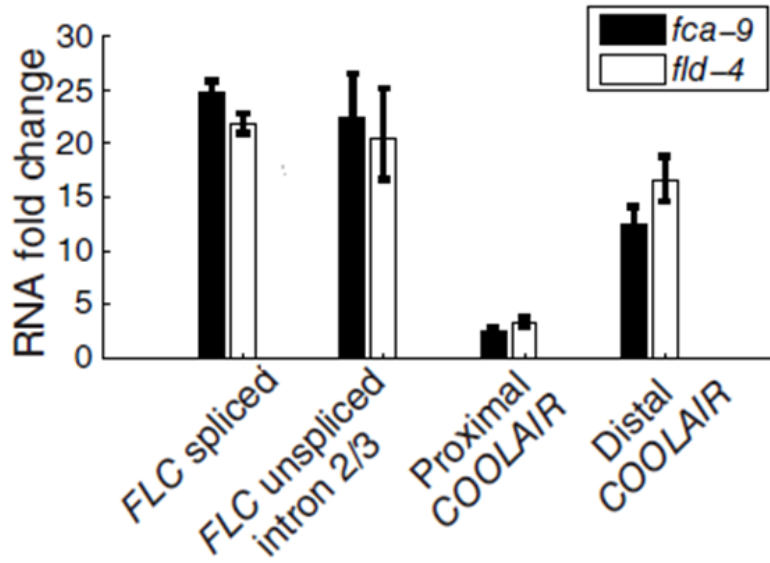


Figure 2.1: RNA fold upregulation in *fca-9* mutant compared to Col: spliced and unspliced *FLC* (~ 25x), proximal (~ 2x) and distal *COOLAIR* (~ 13x). Also shown is the upregulation in the *fld-4* mutant. Values are mean $\pm$ s.e.m. from 3 to 6 independent samples. Experiments performed by Zhe Wu.

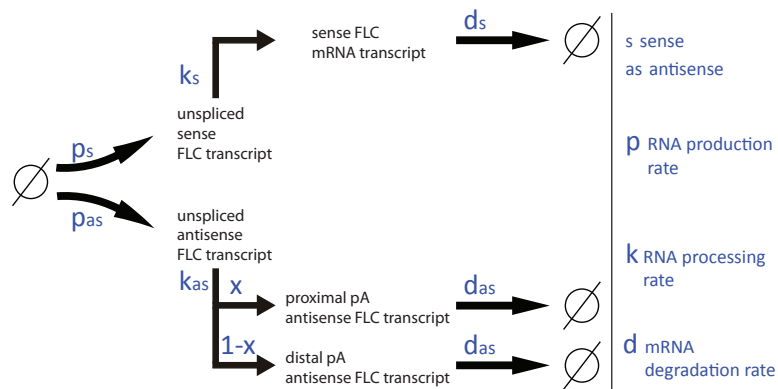


Figure 2.2: Overview of reactions for a simple ODE model (Equation 2.1) of autonomous pathway regulation by FCA and FLD of *FLC*. The model assumes that the rates  $p_s$  and  $p_{as}$  are increased and  $x$ , the probability to choose the proximal poly(A) site, reduced in *fca-9* and *fld-4* mutants as compared to Col. The other parameters are not genotype specific.

ing  $(k_s, k_{as})$  and mature RNA degradation  $(d_s, d_{as})$ . Furthermore we have probability  $x$  ( $1-x$ ) to choose the proximal (distal) splice/ polyadenylation (poly(A)) site. This results in:

$$\begin{aligned}
 \frac{dU_s(t)}{dt} &= p_s - k_s U_s \\
 \frac{dR_s(t)}{dt} &= k_s U_s - d_s R_s \\
 \frac{dU_{as}(t)}{dt} &= p_{as} - k_{as} U_{as} \\
 \frac{dR_{as}^p(t)}{dt} &= x k_{as} U_{as} - d_{as} R_{as}^p \\
 \frac{dR_{as}^d(t)}{dt} &= (1-x) k_{as} U_{as} - d_{as} R_{as}^d.
 \end{aligned} \tag{2.1}$$

We now assume the RNA levels are in steady state (time derivative of RNA species equals zero). Then the solutions to Equation 2.1 are:

$$\begin{aligned}
 U_s &= \frac{p_s}{k_s} \\
 R_s &= \frac{p_s}{d_s} \\
 U_{as} &= \frac{p_{as}}{k_{as}} \\
 R_{as}^p &= x \frac{p_{as}}{d_{as}} \\
 R_{as}^d &= (1-x) \frac{p_{as}}{d_{as}}.
 \end{aligned} \tag{2.2}$$

To relate the model to the experimental measurements we note that the measurements are fold changes of *fca-9* (*fld-4*) over Col. In the model this fold change  $\mathbb{F}$  is the ratio of levels in the mutant over Col. In line with the hypothesis that RNA production is regulated by the autonomous pathway, we now assume that the rates  $p_s$  and  $p_{as}$  are genotype specific. Furthermore the kinetic coupling hypothesis suggests that  $x$  is decreased in *fca-9* (or *fld-4*) mutants:  $x^{fca-9} < x^{Col}$ . We

then find that the model fold changes result in:

$$\begin{aligned}
 \mathbb{F}(U_s) &= \frac{p_s^{fca-9}}{p_s^{Col}} \\
 \mathbb{F}(R_s) &= \frac{p_s^{fca-9}}{p_s^{Col}} \\
 \mathbb{F}(U_{as}) &= \frac{p_{as}^{fca-9}}{p_{as}^{Col}} \\
 \mathbb{F}(R_{as}^p) &= \frac{x^{fca-9}}{x^{Col}} \frac{p_{as}^{fca-9}}{p_{as}^{Col}} \\
 \mathbb{F}(R_{as}^d) &= \frac{(1 - x^{fca-9})}{(1 - x^{Col})} \frac{p_{as}^{fca-9}}{p_{as}^{Col}}.
 \end{aligned} \tag{2.3}$$

This model thus explains why unspliced and spliced *FLC* upregulation is similar: because the RNA production rate increase affects both species similarly. For instance a production rate increase  $\frac{p_s^{fca-9}}{p_s^{Col}} = 25x$  between Col and *fca-9* would be able to fit the data shown in Figure 2.1. Furthermore this model indeed also explains that  $\mathbb{F}(R_{as}^d) > \mathbb{F}(R_{as}^p)$  due to the kinetic coupling hypothesis. As long as antisense production is also upregulated  $\frac{p_{as}^{fca-9}}{p_{as}^{Col}} > 1$ , then there should be a concomittant value for the splice/poly(A) site probability decrease  $\frac{x^{fca-9}}{x^{Col}} < 1$  that could then explain the differential proximal and distal fold upregulation observed in Figure 2.1, for instance  $\frac{p_{as}^{fca-9}}{p_{as}^{Col}} = 6$  and  $\frac{x^{fca-9}}{x^{Col}} = \frac{1}{3}$  in case of  $x^{Col} = \frac{7}{11}$ . We conclude that the RNA fold changes in autonomous pathway mutants are consistent with differential RNA production upregulation and a kinetic coupling hypothesis between antisense elongation and splicing/ polyadenylation [98, 127].

### 2.2.2 Effects of initiation, elongation and termination on Pol II levels

To investigate the autonomous pathway regulation in more mechanistic detail, we now focus on *FLC* transcription by Pol II. Recently, Ehrensberger et al. [204] analysed mathematically the Pol II levels measured in chromatin immunoprecipitation (ChIP) experiments. Here we extend this analysis and include the effects of termination on Pol II levels across a gene. For simplicity, we first develop our theory for sense transcription, before generalising to include both sense and antisense in subsection 2.2.4.

We define the initiation rate ( $F$ ) as the number of Pol II per unit time that successfully bind to the transcription start site (TSS) and become competent to elongate, with velocity  $v$ . Since this elongation velocity is commonly referred to as elongation rate, we use the latter terminology throughout this thesis, whilst bearing in mind that  $v$  indicates a velocity (units:  $bp/s$ ) and not a rate (units:  $s^{-1}$ ). Processes such as unsuccessful Pol II initiation, formation of a transcription elongation complex, promoter proximal pausing, arrest or slow elongation near the TSS (thereby inhibiting new initiation) also affect the magnitude of the initiation rate [204]. Nevertheless our analysis here is general enough to account for these situations. The intrinsic Pol II processivity is generally believed to be high [204], therefore we first analyse a situation where Pol II continues to elongate until it is terminated by some active mechanism. Usually termination occurs after Pol II reaches a poly(A) site, as a consequence of cleavage / polyadenylation of the nascent RNA transcript and subsequent eviction by a 'torpedo' mechanism, mediated by the 5'-3' exonuclease Xrn [47]. Even though this involves multiple steps, we assume in our minimal modelling approach that Pol II termination and RNA 3' processing occur in one step with a probability per unit time of  $k_{pA}$ .

If there is a single termination site, the density  $P_{II}$  of (elongating) Pol II at a distance  $x$  from the TSS, upstream of the termination site ( $3'$ ), is specified by the following equation:

$$\frac{\partial P_{II}(x, t)}{\partial t} = F\delta(x) - \frac{\partial}{\partial x} [v(x)P_{II}(x, t)].$$

Here  $\delta(x)$  denotes the Dirac delta function. At steady state and downstream of the TSS, this equation is solved by:

$$P_{II}(x) = \begin{cases} \frac{F}{v(x)} & \text{if } 0 \leq x < 3', \\ 0 & \text{if } x < 0. \end{cases} \quad (2.4)$$

as found in [204]. Intuitively, a higher initiation rate will increase the Pol II density, whereas a faster elongation rate will reduce it, consistent with Equation 2.4. Pol II acceleration during early elongation [205], promoter proximal pausing [38, 39, 206], transcriptional pausing, backtracking and arrest [36, 207] can all be viewed as instances of a spatially varying elongation rate  $v(x)$ .



Next we consider the situation where Pol II terminates whilst pausing near a poly(A) site [47]. The flow (units:  $s^{-1}$ ) of Pol II through position  $x$  is given by  $v(x)P_{II}(x, t)$ . So in steady state the flow into the 3' termination site is equal to  $F$ . The rate of terminating Pol II is determined by the rate  $k_{pA}$  of RNA 3' processing/Pol II termination as introduced above. Since in steady state the flows balance, we have for the number of Pol II,  $N_{PII}$ , (not density in this case) at the termination site:  $F = k_{pA}N_{PII}(3')$  and thus:  $N_{PII}(3') = \frac{F}{k_{pA}}$ . Hence, the density of Pol II at that position is given by :

$$P_{II}(3') = \frac{F}{L_{PII}k_{pA}}, \quad (2.5)$$

where  $L_{PII}$  is the length of the termination site. If Pol II termination is sufficiently slow, this analysis can explain why 3' Pol II levels can be higher than those in the gene body (see below).

We emphasize that in this model every Pol II that reaches the 3' termination site has produced one unspliced transcript. In this case the (steady state) RNA production rate  $p$  is determined by the flow of Pol II terminating at the 3' site, which is equal to  $F$  and thus generally independent of the elongation rate  $v$ . We do not suggest that  $p$  is always independent of elongation. Slow elongation causes an upper limit on the initiation rate due to excluded volume effects, more commonly known as traffic jams [204]. However, even in this case the production rate equals the initiation rate. We conclude that in the presence of one 3' Pol II termination site, the steady state RNA production rate  $p$  equals the initiation rate  $F$ .

### 2.2.3 *RNA fold changes do not reflect Pol II occupancy changes*

In the previous sections we concluded that a 25x increase in Pol II initiation in an *fca-9* or *fld-4* mutant would produce 25x more unspliced RNA, spliced RNA and Pol II levels along *FLC* (Equation 2.3, Figure 2.3A). This assumes that the *FLC* degradation, splicing/3' processing rates, Pol II processivity, elongation are all unaffected. In contrast to these expectations, both total Pol II and productively elongating Pol II (Ser2-P, see subsection 1.1.1 for description) showed small

changes (2 – 3x) across the whole gene between the two expression states (Figure 2.3B,C).

Importantly, measurements on a highly expressed gene (*ACT7*), and a Pol IV/V transcribed region (*IGN5*), showed that a wide dynamic range ( $> 1000\times$ ) could be detected in the Pol II ChIP assay (Figure 2.3B,C). In addition, Pol II levels at *FLC* were well above background at *IGN5* (Figure 2.3B,C; [208]). Moreover, specific dilutions of *FLC* chromatin, without changing the overall amount of chromatin, resulted in an approximately linear reduction in the Pol II ChIP signal at *FLC* (data not shown). This implies approximate linearity between the Pol II ChIP signal and the Pol II concentration at *FLC*. Therefore, the fold changes measured in our ChIP assay are very unlikely to underestimate the order of magnitude of the changed Pol II occupancy. Cell-specific *FLC* expression variation is also highly unlikely to underlie this difference in RNA and Pol II upregulation, as both assays use whole plant seedlings and thus reflect population averages. We also did not experimentally observe spatially varying levels of Pol II over the gene body, in either the Col, *fca-9* or *fld-4* cases (Figure 2.3B,C). As such variation could occur with spatially varying elongation rates, we therefore now proceed with our analysis setting  $\nu$  constant. Based on the above observations, we conclude that regulation of sense *FLC* transcription is unlikely to occur solely through changes in transcriptional initiation.

#### 2.2.4 *FLC transcriptional dynamics can be explained by coordination of initiation and elongation.*

To further understand how FCA and FLD mediate *FLC* repression at a transcriptional level, we used the above developed analytical relations between mRNA production and Pol II initiation, elongation and termination to model the transcriptional dynamics at *FLC* (Figure 2.4). The experimental data described above (Figure 2.1, Figure 2.3) was used to parameterize the model. Since Pol II ChIP is not strand specific, we now also include antisense transcription in our analytical model fits to the Pol II data. Below the subscript 's' refers to sense, and subscript 'as' to antisense.

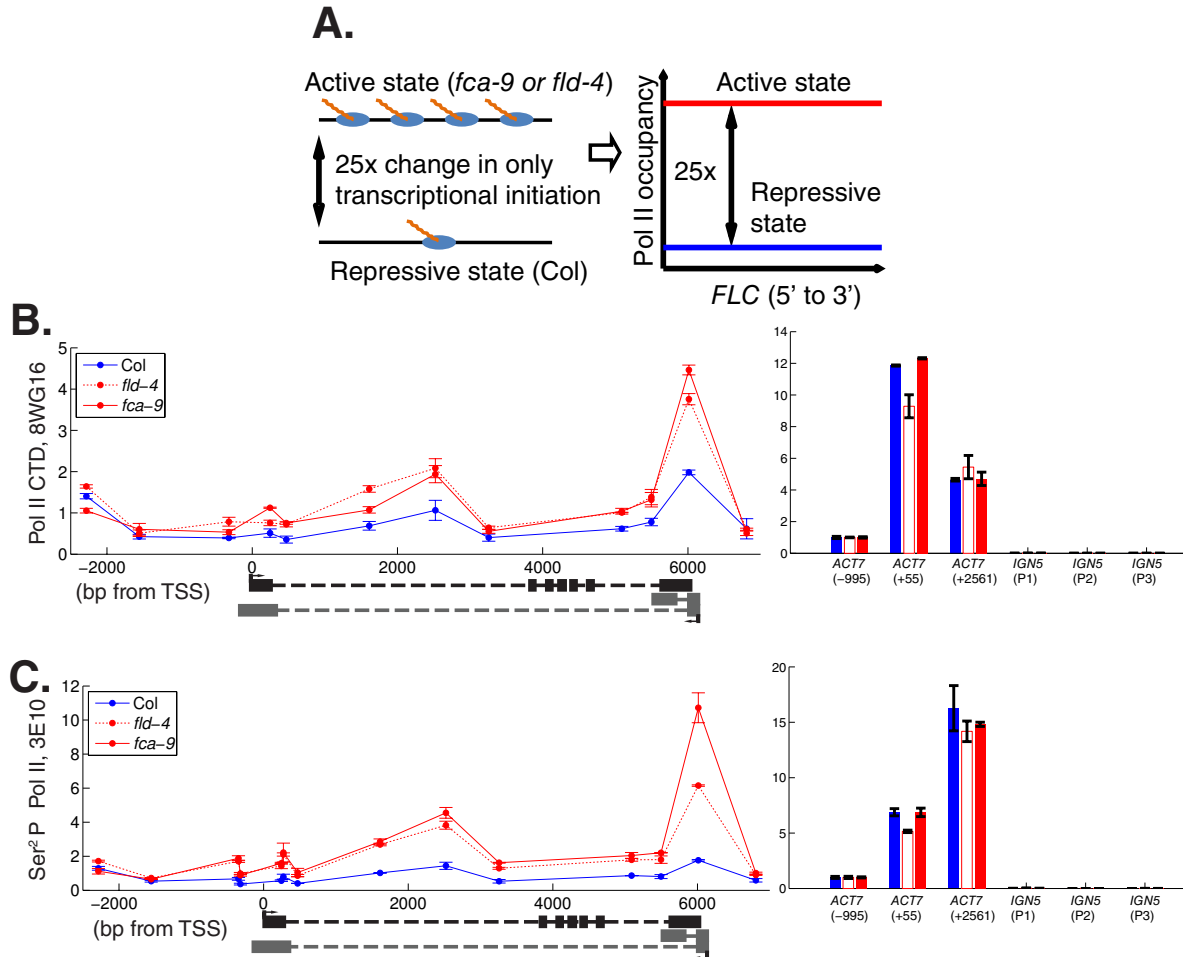


Figure 2.3: (A) Schematic illustration of a scenario where transcriptional initiation is the only difference between Col and *fca-9*. In this case, a ~ 25x fold change in Pol II occupancy should be observed as illustrated on the right. (B-C) ChIP experiments assaying Pol II occupancy across *FLC* using the antibodies for total Pol II anti-CTD 8WG16 (B) and for productively elongation Pol II anti-Ser2 P CTD 3E10 (C). The bar charts at the bottom indicate Pol II levels at various control genes. Levels at *IGN5* are measured by three over-lapping primer pairs (P1-P3). Values are mean $\pm$ s.e.m. from 2 independent samples, with data presented as the ratio of Pol II at *FLC* / input at *FLC* to Pol II at *ACT7* (-995) / input at *ACT7* (-995). Experiments performed by Zhe Wu.

The analytical model has the following (*a priori* unknown) parameters: four initiation rates  $F_s^{\text{Col}}$ ,  $F_s^{\text{fca-9}}$ ,  $F_{as}^{\text{Col}}$ ,  $F_{as}^{\text{fca-9}}$ , two elongation rates (not strand specific):  $v^{\text{Col}}$ ,  $v^{\text{fca-9}}$ , the sense termination rate  $k_{pA,s}$ , the proximal antisense termination rate  $k_{pA,prox}$ , and the distal antisense termination rate  $k_{pA,dist}$ . The *FLC* locus is discretized with a grid size  $dx$  specified by the Pol II footprint  $L_{\text{PII}} = 30bp$ , with site number 0 to 208. In general total Pol II levels at position  $x$ ,  $\text{PII}(x)$ , are a sum of sense and antisense transcribing Pol II levels. Below we mostly drop the Col and *fca-9* subscripts for brevity, although the initiation and elongation rates are genotype specific.

Antisense transcribing Pol II can terminate either at the proximal or distal site. Whilst a Pol II resides at the proximal termination site it can either elongate beyond the site with elongation rate  $v$  or terminate with rate  $k_{pA,prox}$ . In this way the rate of termination depends on the time a Pol II resides at the proximal termination site, which itself is dependent on the elongation rate. This mathematically describes the kinetic coupling mechanism hypothesized in [98, 127]. Of the antisense Pol II levels, a fraction  $1 - \exp\left(-\frac{k_{pA,prox}}{v}L_{\text{PII}}\right)$  will have terminated at the proximal poly(A) site ( $x = 186$ ), so that the flow of Pol II through to the distal site will be  $\widetilde{F}_{as} := F_{as} \exp\left(-\frac{k_{pA,prox}}{v}L_{\text{PII}}\right)$ . This result will be more formally derived below in subsection 2.2.7 in the context of the theory on premature Pol II termination in a 'window of opportunity' of 1 site at the proximal poly(A) site (Equation 2.13).

Generalising from subsection 2.2.2 and using the result above on proximal antisense termination, we find that at the antisense distal termination site,  $x = 0$ , the Pol II density is:

$$\text{PII}(0) = \frac{\widetilde{F}_{as}}{L_{\text{PII}}k_{pA,dist}} = \frac{F_{as}}{L_{\text{PII}}k_{pA,dist}} \exp\left(-\frac{k_{pA,prox}}{v}L_{\text{PII}}\right).$$

Between the antisense distal termination site and the sense TSS,  $x = 1 - 3$ , we have:

$$\text{PII}(x) = \frac{\widetilde{F}_{as}}{v} = \frac{F_{as}}{v} \exp\left(-\frac{k_{pA,prox}}{v}L_{\text{PII}}\right).$$

At and between the sense TSS and the antisense proximal termination site,  $x = 4 - 186$ , we have:

$$\text{PII}(x) = \frac{\widetilde{F}_{as}}{v} + \frac{F_s}{v} = \frac{F_{as}}{v} \exp\left(-\frac{k_{pA,prox}}{v}L_{\text{PII}}\right) + \frac{F_s}{v}.$$

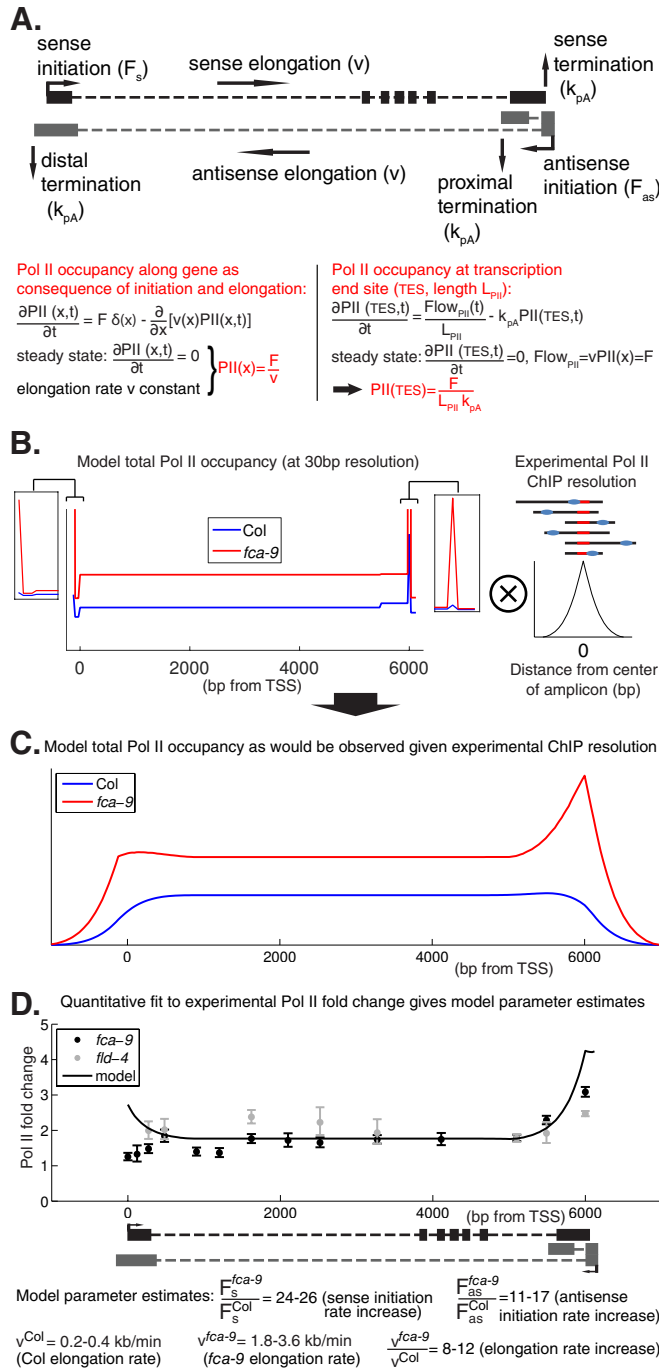


Figure 2.4: (A) Outline of the analytic mathematical model for *FLC* transcription. We used the annotated RNA 3' ends to determine where Pol II could drop off the template after elongation. (B) Total Pol II levels in Col and *fca-9* across *FLC* from the model at a resolution of 30 bp. The *fld-4* mutant model results are identical to *fca-9*. Also shown on the right is a schematic of the convolution process with experimental Pol II ChIP fragment size distribution. (C) Total Pol II levels in Col and *fca-9* across *FLC* from the model convolved with experimental Pol II ChIP fragment size distribution. (D) Experimental and model Pol II fold up-regulation. Experimental values are mean  $\pm$  s.e.m. from 2 to 5 independent samples, including data shown in Figure 2.3. Model fold changes are ratio of profiles shown in (C). Experiments performed by Zhe Wu.

Between the antisense proximal termination site and the sense termination site,  $x = 187 - 203$ , we have:

$$\text{PII}(x) = \frac{F_{as}}{\nu} + \frac{F_s}{\nu}.$$

At the sense termination site,  $x = 204$ , we have:

$$\text{PII}(204) = \frac{F_{as}}{\nu} + \frac{F_s}{L_{\text{PII}}k_{pA,s}}.$$

Finally, beyond the sense termination site,  $x = 205 - 208$ , we have:

$$\text{PII}(x) = \frac{F_{as}}{\nu}.$$

To directly compare the model Pol II distribution with our experimental data, we convoluted model Pol II levels for both Col and *fca-9* with the experimental fragment size distribution (Figure 2.4B, see subsection 2.4.1 and subsection 2.4.2 for more details). The shape of the convoluted Pol II profile  $C(x)$  across the gene (Figure 2.4C) can be compared to the experimental Pol II data in Figure 2.3, while the ratio of convoluted Pol II levels can be directly compared to the experimental Pol II fold upregulation (Figure 2.4D). Note that we do not include interference between sense and antisense transcription. This simplification is justified in the case of low absolute initiation rates and such an assumption is reasonable as transcriptional interference is incompatible with the observed positive correlation between sense and antisense RNA transcripts (Figure 2.1). We will revisit transcriptional interference effects in later sections. Elevated Pol II levels at the 3' of *FLC* resulted from sense termination and proximal antisense transcription (Figure 2.4).

As explained in subsection 2.2.2, spliced sense *FLC* mRNA upregulation is determined by the ratio  $\frac{F_s^{fca-9}}{F_s^{\text{Col}}}$ . Hence this ratio is used in the fit in Figure 2.5. To fit unspliced sense *FLC*, we note that the forward and reverse primers lie in intron2 and 3 respectively. Thus, signal can only be picked up when intron2 and the majority of intron3 have already been transcribed. Due to the small sizes involved, we can take as a simple proxy for this species the transcripts that contain intron2. This process is mathematically well described by the ODE for  $U_s$  (see Equation 2.1) that gives rise to unspliced sense *FLC* levels of  $\frac{F}{k_s}$  (Equation 2.2). The fold upregulation of unspliced RNA

will then be  $\frac{F_s^{fca-9}}{k_s} / \frac{F_s^{Col}}{k_s} = \frac{F_s^{fca-9}}{F_s^{Col}}$  under the assumption that the splicing rate does not change (Equation 2.2). For the distal antisense RNA, levels will be proportional to both the antisense initiation rate and to the fraction of antisense RNA for which the proximal poly(A) site is not utilised, but will be inversely proportional to the distal antisense degradation rate. Assuming that the transcript lifetime does not change between genotypes, we find for the fold change:

$$\mathbb{F}(distal) = \frac{F_{as}^{fca-9}}{F_{as}^{Col}} \frac{\exp\left(-\frac{k_{pA,prox}}{v^{fca-9}} L_{PII}\right)}{\exp\left(-\frac{k_{pA,prox}}{v^{Col}} L_{PII}\right)}. \quad (2.6)$$

Performing a similar analysis for the proximal antisense, we find:

$$\mathbb{F}(proximal) = \frac{F_{as}^{fca-9}}{F_{as}^{Col}} \frac{1 - \exp\left(-\frac{k_{pA,prox}}{v^{fca-9}} L_{PII}\right)}{1 - \exp\left(-\frac{k_{pA,prox}}{v^{Col}} L_{PII}\right)}. \quad (2.7)$$

From the literature [41], we fix the value of  $k_{pA,s} = k_{pA,prox} = \frac{1}{50} s^{-1}$ . We then allowed  $k_{pA,distal}$  to deviate at most five fold from  $k_{pA,s}$ .

Since Pol II levels are measured in our experiments as % input and are thus relative, absolute initiation rates cannot be directly inferred from Pol II ChIP using our analytical model. In our parameter fitting methodology, we handled this issue as follows: we first set  $F_s^{Col}$  to  $0.1 min^{-1}$  and determined other parameters relative to this arbitrary value. We then performed a parameter sweep for the following variables to see which values were able to fit the data:  $\frac{F_s^{fca-9}}{F_s^{Col}}$ ,  $\frac{F_{as}^{Col}}{F_s^{Col}}$ ,  $\frac{F_{as}^{fca-9}}{F_{as}^{Col}}$ ,  $v^{Col}$ ,  $\frac{v^{fca-9}}{v^{Col}}$ ,  $\frac{k_{pA,distal}}{k_{pA,s}}$ . We also verified that our choice of  $F_s^{Col} = 0.1 min^{-1}$  was indeed arbitrary and that other choices led to the same results. The following criteria all needed to be met for a set of parameters to be considered as being able to fit the data:

- $\log\left(\frac{F_s^{fca-9}}{F_s^{Col}}\right)$  within mean $\pm$ s.e.m. of experimental log-fold change of spliced *FLC*.
- $\log(1.4) \leq \log(\mathbb{F}(proximal)) \leq$  mean+s.e.m. of experimental log-fold change of proximal *COOLAIR*.
- mean-s.e.m. of experimental log-fold change of distal *COOLAIR*  $\leq \log(\mathbb{F}(distal)) \leq \log(19)$ .

- No 5' Pol II peak observable in *fca-9* despite frequent distal antisense termination, using convoluted Pol II levels:

$$\frac{C^{fca-9}(x=0)}{C^{Col}(x=100)} \leq 1.1$$

- For comparison of gene body Pol II upregulation, using convoluted Pol II levels:

$\log\left(\frac{C^{fca-9}(x=113)}{C^{Col}(x=113)}\right)$  within mean±s.e.m. of experimental Pol II log-fold change at corresponding primer ( $x = 113 \times 30 = 3390bp$  from TSS) in Figure 2.4D.

- Limited 3' Pol II fold change, where, using convoluted Pol II levels:

mean–s.e.m. of experimental log-fold upregulation at corresponding primer  $\leq \log\left(\frac{C^{fca-9}(x=204)}{C^{Col}(x=204)}\right) \leq \log(4.5)$ .

Using this procedure, we found that parameters values within the following ranges could fit our data:

$$\frac{F_s^{fca-9}}{F_s^{Col}} = 24 - 26,$$

$$\frac{F_{as}^{Col}}{F_s^{Col}} = 2 - 5,$$

$$\frac{F_{as}^{fca-9}}{F_{as}^{Col}} = 11 - 17,$$

$$v^{Col} = 0.2 - 0.4 \text{ kb/min},$$

$$\frac{v^{fca-9}}{v^{Col}} = 8 - 12,$$

$$\frac{k_{pA,distal}}{k_{pA,s}} = 2 - 5.$$

Note that using this methodology, after fixing  $k_{pA,s}$ , the absolute elongation rate is determined (between 0.2 – 0.4 *kb/min* for Col), as well as the fold change between *fca-9* and Col (8 – 12 fold). This method suggests that in Col the antisense initiates more frequently than the sense. The values used for the fits in Figure 2.4, Figure 2.5, Figure 2.6 and Figure 2.7 are:  $\frac{F_s^{fca-9}}{F_s^{Col}} = 25$ ,  $\frac{F_{as}^{Col}}{F_s^{Col}} = 4$ ,  $\frac{F_{as}^{fca-9}}{F_{as}^{Col}} = 14$ ,  $v^{Col} = 0.3 \text{ kb/min}$ ,  $\frac{v^{fca-9}}{v^{Col}} = 10$ ,  $\frac{k_{pA,distal}}{k_{pA,s}} = 5$ . This results in  $\mathbb{F}(\text{proximal}) = 1.5$  and  $\mathbb{F}(\text{distal}) = 15.6$  (Figure 2.5). Hence, our analytic model was able to generate a Pol II profile along *FLC* (Figure 2.4B-D) similar to the experimental Pol II ChIP data (Figure 2.3B,C). At the same time, this model also reproduced the *FLC* spliced, unspliced and



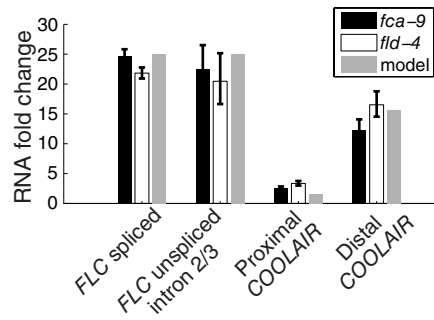


Figure 2.5: RNA fold upregulation in *fca-9* and *fld-4* mutants compared to Col: experimental data as in Figure 2.1. The model values are the fits to the experimental data.

*COOLAIR* fold upregulation in *fca-9* and *fld-4* (Figure 2.5). Intuitively, the small increase of Pol II ChIP signal in the transcriptionally active *fca-9* and *fld-4* mutants could be explained by a coordinated increase in initiation and elongation rates. To quantitatively fit the approximately two-fold Pol II occupancy increase (Figure 2.4D) an 8 – 12x fold faster elongation during transcription in the *fca-9* and *fld-4* mutants as compared to Col was required.

The relative increase in elongation rates between *fca-9* and Col can also be obtained in a more intuitive way. Suppose that sense initiation predominates across *FLC*. In that case, the required elongation rate change would have to be approximately  $25/1.8 = 14x$ , the sense fold upregulation over the Pol II fold change. A second estimate for elongation rate changes can be obtained by assuming predominant antisense initiation. In this case, the required elongation rate change would have to be  $12.5/1.8 = 7x$ , the distal antisense fold upregulation over the Pol II fold change. Therefore a range of 7 – 14x for the elongation rate change is a model-independent estimate for the range of elongation rate changes that are consistent with our data, and is consistent with our simulation fitting procedure above. Furthermore, using an experimentally determined value for the termination rate  $\frac{1}{50}s^{-1}$  [41], absolute elongation rates could be inferred from the model, yielding 0.2 – 0.4 kb/min (Col) and 1.8 – 3.6 kb/min (*fca-9* and *fld-4*). We conclude that the experimental observations described above, can be explained by a model of transcriptional dynam-

ics that assumes a coordinated increase of the initiation and elongation rates.

#### 2.2.5 *Initiation and elongation rate effects on intronic RNA expression*

Next we aimed to design an experiment to test the predicted coordinate increase in initiation and elongation rates. Measurement of elongation rates on a subset of highly expressed, long mammalian genes (>50 kb) has been achieved using GRO-seq [205]. This is an elongation inhibition based method relying on rapid removal of a transcription inhibitor, which is difficult in whole organisms [209, 210]. An alternative approach based on an MS2 fusion [41, 211] was not successful since the *FLC-MS2* fusion was not expressed at sufficient levels. To overcome these limitations, we investigated theoretically how intronic RNA expression depends on initiation and elongation.

Since nascent RNA is produced from 5' to 3', and is exported from the locus after Pol II has reached the 3' end of the gene, exonic 5' RNA generally resides on the chromatin longer than 3' exonic RNA. This generates a nascent RNA profile with declining levels from the 5' to 3' end [44, 212]. The same principle can be applied to intronic sequences: for RNA close to the intron donor site the time between production and intron cleavage is longer than for RNA close to the intron acceptor site [213]. Therefore, intronic RNA levels are again expected to reduce from the 5' to 3' end of the intron. Elongation rates influence how quickly Pol II reaches a poly(A) or intron acceptor site and thereby affect how quickly RNA export or splicing, and subsequent lariat degradation, can occur after the initiation of transcription. As a result, Pol II elongation rates can thus affect the lifetime of nascent RNA (in contrast to mature RNA discussed in subsection 2.2.2).

To formalize this reasoning, let  $v$  be the (spatially independent) elongation rate,  $k_s = 1/T_s$  be the splicing rate (intron cleavage rate) given that the intron has been fully transcribed, and  $k_x = 1/T_x$  be the RNA export rate from the locus, given that the full length RNA has been

transcribed. Furthermore, we assume that, after splicing, linearized lariat RNA is degraded from 5' to 3', from intron donor site ( $ID$ ) to intron acceptor site ( $IA$ ), with degradation rate  $k_l$  (units:  $bp/s$ ) [214]. Here,  $x$  is again the distance from the TSS.

First we analyse total intronic RNA levels,  $RNA_{total}(x)$ , which are generally made up of three fractions: firstly, Pol II-bound RNA arising from Pol II that has elongated beyond  $x$  but has not yet reached the  $IA$  ( $RNA_i(x)$ ); secondly, RNA with the complete intron transcribed that is competent to be spliced out ( $RNA_s$ ); and thirdly, intronic degradation products in the process of being degraded  $RNA_{lariat}(x)$ . This decomposition leads to

$$RNA_{total}(x) = RNA_i(x) + RNA_s + RNA_{lariat}(x).$$

For  $RNA_i(x)$ , all Pol II present along the intron that has elongated beyond  $x$  will contribute. Hence, we have:

$$RNA_i(x) = \int_x^{IA} PII(y) dy.$$

Overall, we therefore find:

$$\begin{aligned} RNA_{total}(x) &= \int_x^{IA} PII(y) dy + RNA_s + RNA_{lariat}(x) \\ &= F \left[ \frac{IA - x}{v} + T_s + \frac{x - ID}{k_l} \right] \end{aligned}$$

where the expression for  $RNA_{lariat}(x)$  follows similar reasoning as for  $RNA_i(x)$  above. We assume here that degradation begins as soon as splicing has occurred. However, if there is a delay, this can be straightforwardly absorbed into the timescale  $T_s$ . Modelling the donor and acceptor site cleavage steps separately [213] is not essential for purely intronic sequences: the timescale  $T_s$  represents the time after which cleavage of both has occurred. RNA levels across the exon/intron junction at the donor site can be modelled by taking  $x = ID$  and replacing  $T_s$  with a timescale for donor site cleavage:  $T_{ID} \leq T_s$ . We also considered the potential presence of 3' to 5' lariat RNA degradation, with rate  $k_{l,35}$  (units:  $bp/s$ ), in addition to 5' to 3' degradation [214]. In this case we can replace the expression for  $RNA_{lariat}(x)$  with  $F \cdot \min\left(\frac{x-ID}{k_i}, \frac{IA-x}{k_{l,35}}\right)$ .

Note that we expect only  $F$  and  $\nu$  to alter between Col and the *fca-9* mutant, as we concluded in the previous section. Total intronic RNA fold changes  $\mathbb{F}$  can be experimentally measured in total RNA assays:

$$\mathbb{F}(\text{RNA}_{\text{total}}(x)) = \frac{F^{\text{fca-9}} \left[ \frac{IA-x}{\nu^{\text{fca-9}}} + T_s + \frac{x-ID}{k_i} \right]}{F^{\text{Col}} \left[ \frac{IA-x}{\nu^{\text{Col}}} + T_s + \frac{x-ID}{k_i} \right]}. \quad (2.8)$$

Furthermore, we see that, in the presence of elongation rate changes, the intronic RNA fold change profile as a function of position will be qualitatively different in the absence of intronic degradation products ( $\text{RNA}_{\text{lariat}}$ ):

$$\mathbb{F}(\text{RNA}_{\text{nasc}}(x)) = \frac{F^{\text{fca-9}} \left[ \frac{IA-x}{\nu^{\text{fca-9}}} + T_s \right]}{F^{\text{Col}} \left[ \frac{IA-x}{\nu^{\text{Col}}} + T_s \right]}. \quad (2.9)$$

where we assume that the splicing time  $T_s$  is much shorter than the combined timescale to elongate across the locus and then be exported. Experimentally, splicing has been shown to take on the order of 100s [44].

The  $\text{RNA}_{\text{lariat}}$  fraction, found to be nucleoplasmic but not chromatin-bound [212], is likely to be largely washed away in a chromatin-bound RNA assay. Close to the intron 3' end, this analysis indicates that nascent RNA levels will be mostly determined by the ratio of the initiation rate to the splicing rate, both of which are independent of the elongation rate. Further away transcripts from Pol II still traversing the intron will also determine RNA levels, and hence the ratio of the initiation rate to the elongation rate will become important. Due to this spatially varying dependence on the elongation rate, which is predicted to be 8 – 12x faster in the active (*fca-9*) versus repressed (Col) state, we therefore predict a spatially varying fold upregulation of nascent RNA along *FLC* intron1.

Lastly, the model predicts that exonic, chromatin-bound RNA levels,  $\text{RNA}_{\text{exon}}(x)$ , are also dependent on elongation rates, and are determined by two populations: firstly, RNA bound to Pol II, where the Pol II has elongated beyond  $x$  but not yet reached the *FLC* 3' end, and secondly RNA that is full length and residing at the locus prior

to export ( $RNA_x$ ):

$$\begin{aligned} RNA_{exon}(x) &= \int_x^{3'} PII(y) dy + RNA_x \\ &= F \left[ \frac{3' - x}{v} + T_x \right]. \end{aligned} \quad (2.10)$$

### 2.2.6 Coordinated initiation and elongation generates RNA upregulation patterns along *FLC* intron1.

We next tested the key model prediction of the spatially varying fold upregulation of nascent RNA (Figure 2.6A,B) by measuring the chromatin-bound RNA profile at *FLC* (Figure 2.6C). Since measurements were performed on *FLC* intron1 we now use *I1A* and *I1D* to indicate the respective intron1 acceptor and donor sites in our model fold changes Equation 2.8 and Equation 2.9. In the first kb of intron1, as predicted by the model, there was only a small fold increase in *fca-9* as compared to Col (Figure 2.6C; Figure 2.7). This is due to the dependence on the ratios of the initiation and elongation rates and their coordinated increases in *fca-9* (Figure 2.6B). By contrast, the fold upregulation was much larger close to the intron acceptor site in *fca-9*. This is in agreement with the model Equation 2.9, where we used the experimentally determined splicing rate of  $k_s = \frac{1}{100} s^{-1}$  [44] for both Col and *fca-9*, with other parameters determined from our prior fitting to the Pol II ChIP data (see subsection 2.2.4).

Comparing *fca-9* to Col, the fold upregulation inside exon1 was much larger than at the exon1-intron1 junction (Figure 2.7). This indicates that splicing of intron1 occurs mostly co-transcriptionally, because otherwise a similar fold change for exon and intron signals would be expected. We successfully tested in experiments the model prediction that the chromatin-bound RNA fold change at exon1 for *fca-9* and *fld-4* is significantly larger than the fold change at the exon1/intron1 junction (Figure 2.7), where we used the experimentally constrained RNA export rate  $k_x = \frac{1}{T_x} = 1.7 \cdot 10^{-3} s^{-1}$  [41] in (Equation 2.10).

We also investigated what elongation rate changes between Col and *fca-9* are consistent with the chromatin-bound RNA data independently of any prior knowledge of splicing and elongation rate values.

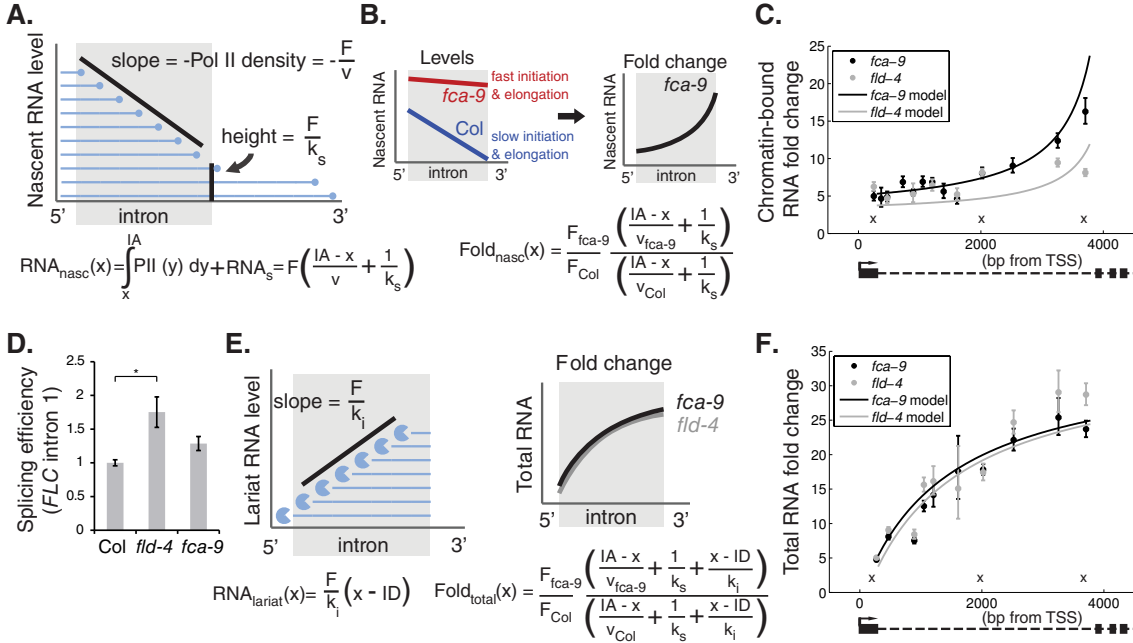


Figure 2.6: Combination of increased initiation and elongation, with co-transcriptional splicing and lariat degradation, leads to distinct RNA profiles along *FLC* intron1. (A) Schematic indicating intronic nascent RNA from Pol II (blue circles) elongating through the intron. The downward slope of nascent RNA levels along the intron is determined by the ratio  $F/\nu$  of initiation rate ( $F$ ) over elongation rate ( $\nu$ ). Abundance of unspliced full-length intron RNAs is given by the ratio of initiation rate over splicing rate ( $k_s$ ). Analytic expression for intronic nascent RNA is shown at the bottom of the figure. (B) Schematic (left panel) indicating model profiles of nascent RNA along *FLC* intron1 in *fca-9* and Col. Between *fca-9* and Col,  $F$  and  $\nu$  are coordinately increased, but with the same splicing rate ( $k_s$ ). These two profiles generate a characteristic pattern of nascent RNA fold changes between *fca-9* and Col (right panel) with analytic expression for intronic nascent RNA fold changes shown at the bottom of the figure. (C) Modelled and experimentally measured chromatin-bound RNA fold changes along *FLC* intron1. The lower increase towards the 3' end in *fld-4* is due to increased splicing rate ( $k_s$ ) as shown experimentally in (D). Crosses indicate positions where data are from 3 different, overlapping primer sets that each show similar results. (D) Estimate of *FLC* intron1 splicing efficiency (intron cleavage rate) in *fld-4* and *fca-9*, normalized to the level in Col. Values are mean  $\pm$  s.e.m. from 3 independent samples. Asterisks indicate statistical significance: for all the figures in this thesis,  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , two-sided unpaired t-test unless specified otherwise. (E) Schematic showing effect of 5' to 3' intronic RNA degradation on lariat RNA levels. Full-length lariat RNA results from splicing and is degraded with rate  $k_l$ . Ratio  $F/k_l$  gives slope of lariat RNA levels along intron. These degradation intermediates, together with the nascent RNA described in (A), make up total intronic RNA. Fold upregulation then generates the characteristic profiles shown. Also shown are analytic expressions for lariat RNA levels and total intronic RNA fold changes. (F) Modelled and experimentally measured total RNA fold changes along *FLC* intron1. (C and F) Experimental values are mean  $\pm$  s.e.m. from at least 3 independent samples. Experiments performed by Zhe Wu.

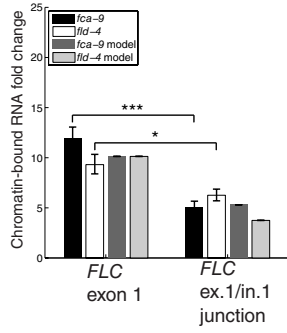


Figure 2.7: Model and experimentally measured chromatin-bound RNA fold upregulation in *fca-9* and *fld-4* compared to Col at *FLC* exon1 and exon1-intron1 junctions. The related mathematical analysis can be found in subsection 2.2.5. Experiments performed by Zhe Wu.

First, note that Equation 2.9 can be written as:

$$\mathbb{F}(RNA_{nasc}(x)) = \frac{F_{as}^{fca-9} \frac{11A-x}{z^{Col} v_{fold}} + 1}{F_{as}^{Col} \frac{11A-x}{z^{Col}} + 1}. \quad (2.11)$$

Here we have replaced the elongation and splicing rates by two different variables:  $z^{Col} = v^{Col}/k_s$  and  $v_{fold} = v^{fca-9}/v^{Col}$ . Note that  $v_{fold}$ , but not  $z^{Col}$ , is the variable that contains information about elongation rate fold changes between Col and *fca-9*. As the ratio  $\frac{F_{as}^{fca-9}}{F_{as}^{Col}}$  is fixed by the spliced sense *FLC* mRNA upregulation,  $z^{Col}$  and  $v_{fold}$  are the only a priori unknowns in Equation 2.11. We were therefore able to perform a nonlinear curve fitting procedure to obtain estimates for these unknown parameters directly from the chromatin-bound RNA fold changes in *fca-9* compared to Col. We log-transformed both the data and Equation 2.11 so that the experimental s.e.m. (value:  $SEM(x)$ ) on the log-fold change could be used to assign the appropriate weight to each data point:  $weight(x) = \frac{1}{SEM^2(x)}$ . With this weight function our method is essentially a  $\chi^2$  goodness of fit method that is commonly used to estimate parameters from data. Indeed Equation 2.11 can fit well ( $R^2 = 0.89$ , F statistic:  $p = 3 \cdot 10^{-14}$ ) to our experimental chromatin-bound RNA fold changes, with estimates for  $z^{Col} = 550 \pm 140$  bp (mean  $\pm$  s.e.m.,  $p = 2.1 \cdot 10^{-3}$ ) and  $v_{fold} = 9.8 \pm 3.8$  ( $p = 2.6 \cdot 10^{-2}$ ). This analysis shows independently of the Pol II data and any potential uncertainty in splicing rates, that to

explain the chromatin-bound RNA upregulation in *fca-9*, elongation rate changes of 6 – 14 fold are required. Thus this analysis provides additional evidence for the elongation rate changes. The procedure also gives an estimate for  $\frac{v^{\text{Col}}}{k_s}$ , where using  $k_s = 0.01s^{-1}$ , we find  $v^{\text{Col}} = 0.3 \pm 0.1 \text{ kb/min}$ . This result gives further confidence in our earlier estimate for  $v^{\text{Col}}$ .

Interestingly, we observed less increase in fold upregulation towards the 3' end of intron1 in *fld-4* as compared to *fca-9* (Figure 2.6C). Given the fold change close to an intron acceptor site is more sensitive to splicing rather than elongation rate changes (Figure 2.6B), we examined if a splicing rate change specific to *fld-4* could explain its differential fold upregulation pattern from *fca-9*. Indeed, we found that we could fit the *fld-4* profile in our model by incorporating a 2 fold faster splicing rate ( $k_s = \frac{1}{50}s^{-1}$ ) in *fld-4* (Figure 2.6C), whilst keeping all other parameters unchanged. We further verified this model prediction of an increased splicing rate in *fld-4* by measuring the splicing efficiency of *FLC* intron1. As predicted, the efficiency was increased 1.8 fold in *fld-4* (Figure 2.6D) but not significantly altered in *fca-9* ( $p = 0.1$ , two-sided unpaired t-test). A simple alternative model with unchanged splicing and elongation rates between Col and *fld-4* would produce a constant chromatin-bound RNA fold-change across intron1. That would be consistent with the chromatin-bound RNA data set in isolation (Figure 2.6C), but implies a change in the initiation rates of only  $\sim 7$  fold, which is inconsistent with our earlier spliced and unspliced *FLC* RNA fold changes (Figure 2.1).

To further support these conclusions we investigated the total intronic RNA profile (Figure 2.6E,F). Such measurements include intron lariat degradation intermediates, which are present in the total but not chromatin-bound RNA fraction (Figure 2.6E) [212]. As detailed in subsection 2.2.5, assuming that lariat degradation occurs from 5' to 3', lariat RNA at the 3' generally exists for longer than that at the 5'. This generates a lariat RNA profile with increasing levels from the 5' to 3' end (Figure 2.6E). Importantly, incorporating this lariat population into the total intronic RNA fold upregulation between *fca-9* and Col (Equation 2.8), without altering the model pa-



parameterization that explained the Pol II and chromatin-bound RNA, produced a predicted profile that is qualitatively different to that found for the chromatin-bound RNA (Figure 2.6B,E). This prediction was also validated experimentally (Figure 2.6F). Compared to the chromatin-bound RNA profile, there was a significantly larger fold increase in the first 2kb of the total intronic RNA profile ( $p = 8 \cdot 10^{-7}$  and  $4 \cdot 10^{-7}$  for *fca-9* and *fld-4* respectively, two-sided Welch's t-test) (Figure 2.6C,F). In the model, we could regenerate such profile (Equation 2.8), by solely incorporating 5' to 3' intron lariat degradation with rate up to  $1.5bp/s$  [214], in line with experimentally determined intron half-lives [44]. Potential additional presence of 3' to 5' degradation [214] with a rate up to  $1bp/s$  did not alter our conclusions. The profiles for total intronic RNA look very similar between *fca-9* and *fld-4* (Figure 2.6F), in contrast to the chromatin-bound data (Figure 2.6C). This similarity is because the lariat RNA effectively extends the half-life of intronic RNA and therefore reduces the effect of the differential splicing rates between *fca-9* and *fld-4* (Figure 2.6F).

Taken together, our total and chromatin-bound intronic RNA profiles provide strong evidence that repression of *FLC* involves a coordinated change of both the initiation and elongation rates. Moreover, the methods we developed here can be used to infer elongation rate changes in whole organisms where pulse-chase experiments are not feasible.

### 2.2.7 Sense premature termination contributes little to *FLC* repression

In this and the following section we describe alternative mathematical models based on premature Pol II termination and co-transcriptional RNA degradation. These were initially investigated to explain the RNA fold upregulations in autonomous pathway mutants Figure 2.1. We then rule out these mechanisms based on the above described data and also further experiments described below.

Previous reports have linked the elongation rate to either Pol II processivity [215] or early termination [53]. In these scenarios, Pol II would terminate prematurely as a result of slow elongation. Our co-

ordination model detailed in the previous sections did not require any such premature termination. However, it is a priori possible that functional RNA production is regulated by the location of Pol II termination. Indeed a high density of promoter-proximal poly(A) site signals reinforce premature termination to limit pervasive transcription in the divergent antisense direction of coding genes [58]. Alternatively, regulation could occur through an effect on Pol II processivity [215]. Several factors are known to influence Pol II processivity *in vivo* in yeast [215, 216]. Accordingly, we now investigate models where premature termination is a key component.

To maintain generality we again do not specify exactly what mechanism(s) constitute early termination. Instead we assume that inside a region  $R$  immediately downstream of the TSS (Figure 2.8A), Pol II can elongate with velocity  $v(x)$  but also terminate early with a rate  $k_t$ . If Pol II elongates beyond the end of region  $R$  we assume it continues to elongate to a definite termination site further downstream, producing a functional RNA. Below, we show that such a model generates a Pol II density that is exponentially decreasing with distance  $x$  within the region  $R$ .

For the above system, we can write down an appropriate equation for the Pol II density  $\text{PII}(x, t)$  at time  $t$  and distance  $x$  from the TSS (within region  $R$ ):

$$\frac{\partial \text{PII}(x, t)}{\partial t} = F\delta(x) - \frac{\partial}{\partial x} [v(x)\text{PII}(x, t)] - k_t \text{PII}(x, t).$$

We next assume that the system is in steady state:  $\frac{\partial \text{PII}(x, t)}{\partial t} = 0$ . Rearranging, we arrive at a homogeneous differential equation for the steady state Pol II density  $\text{PII}(x)$  for  $x > 0$ :

$$\frac{d\text{PII}(x)}{dx} = - \left[ \frac{1}{v(x)} \frac{dv(x)}{dx} + \frac{k_t}{v(x)} \right] \text{PII}(x). \quad (2.12)$$

This equation can be solved analytically giving:

$$\text{PII}(x) = \text{PII}(0) \frac{v(0)}{v(x)} \exp \left( -k_t \int_0^x \frac{dy}{v(y)} \right).$$

To solve for the Pol II density at  $x = 0$ , we can integrate the inhomogeneous steady state equation,

$$0 = F\delta(x) - \frac{d}{dx} [v(x)\text{PII}(x)] - k_t \text{PII}(x),$$

directly in the interval  $[-\varepsilon, \varepsilon]$ , with  $\varepsilon > 0$  and then take the limit  $\varepsilon \rightarrow 0$ :

$$\begin{aligned} 0 &= \lim_{\varepsilon \rightarrow 0} \int_{-\varepsilon}^{\varepsilon} \left[ F\delta(x) - \frac{d}{dx} [v(x)PII(x)] - k_t PII(x) \right] dx \\ &= F - v(0)PII(0). \end{aligned}$$

This leads to  $PII(0) = \frac{F}{v(0)}$  and furthermore the desired Pol II density in the presence of early termination:

$$PII(x) = \begin{cases} 0 & \text{if } x < 0. \\ \frac{F}{v(x)} \exp\left(-k_t \int_0^x \frac{dy}{v(y)}\right) & \text{if } 0 \leq x < R, \\ \frac{F}{v(x)} \exp\left(-k_t \int_0^R \frac{dy}{v(y)}\right) & \text{if } R \leq x < 3', \end{cases} \quad (2.13)$$

In case of antisense termination at the proximal poly(A) site as described in subsection 2.2.4, Equation 2.13 can be used with  $R$  indicating the termination region size, here given by the length of the proximal poly(A) site  $L_{PII}$ ,  $v$  constant and  $k_t = k_{pA,prox}$ .

Generally, in the case of a relatively high termination rate  $k_t$ , this results in a sharp promoter proximal Pol II peak that is potentially indistinguishable in ChIP assays from promoter proximal pausing [204]. In metazoans, during this transcriptional pause, Pol II has been found to either transition into productive elongation or prematurely terminate [39]. Our analysis also predicts that quicker elongation reduces the premature termination probability, as found with Sen1-mediated early termination in yeast [53]. The RNA production rate is affected by the presence of premature termination because not every Pol II that initiates transcription produces a full length transcript. The production rate is still governed by the flow of Pol II through the 3' end poly(A) site. So the RNA production rate is in this model not equal to  $F$  as found in subsection 2.2.2, but rather  $F \exp\left(-k_t \int_0^R \frac{dy}{v(y)}\right)$  in the presence of an premature termination region  $R$  upstream of the respective poly(A) site and can thus be dependent on the elongation rate.

As derived above, premature termination should lead to declining Pol II levels from 5' to 3' in the repressed case (Col) (Equation 2.13, Figure 2.8). However, we found no evidence for this in our Pol II ChIP assay (Figure 2.3B,C). Since total Pol II and serine 2 phosphory-

lated Pol II generated similar upregulations, we also ruled out models based on productive elongation (details not shown) where regulation of the phosphorylation status could affect RNA production [16]. However, since our Pol II ChIP assay is not strand specific, such a sense Pol II decline could in principle be masked by antisense transcription.

To further confirm that premature termination contributes little to *FLC* repression, we performed 3'RACE, a method that determines the sequence of RNA 3' ends, to map the transcripts that end within the promoter-proximal region of *FLC* [208]. We could detect polyadenylated transcripts that terminated within *FLC* intron1, suggesting sense premature termination within intron1 can occur. These transcripts all contained *FLC* exon1 and were mostly alternatively spliced with the same donor site but with a different acceptor site, as compared to the conventional *FLC* intron1 (Figure 2.8B).

Interestingly, we also mapped many transcripts without a poly(A) tail (Figure 2.8). Some are likely splicing intermediates, as their 3' ends correspond to exon 3' ends with the previous introns removed and exons ligated. Other unpolyadenylated transcripts are likely Pol II-bound as their 3' ends mapped to the middle of an exon or intron. Although this technique is not quantitative, among the clones we sequenced (100-150), we saw mostly splicing intermediates in *fca-9*, while other types were more abundant in Col and 35S::FCA, a functional FCA overexpression line in a Col background [114]. This is consistent with slow elongation in these genotypes (Figure 2.8C), because slow elongation would increase the likelihood that a Pol II with the associated nascent RNA would reside within intron1.

To quantify whether premature termination occurs often as compared to production of a full length transcript, we monitored the alternatively spliced intron associated with premature termination (Figure 2.8B). We found these transcripts are in lower abundance compared to unspliced intron1 in Col, *fca-9* and 35S::FCA (Figure 2.8D). If premature termination would be important for FCA-mediated repression, we would expect premature termination to occur relatively

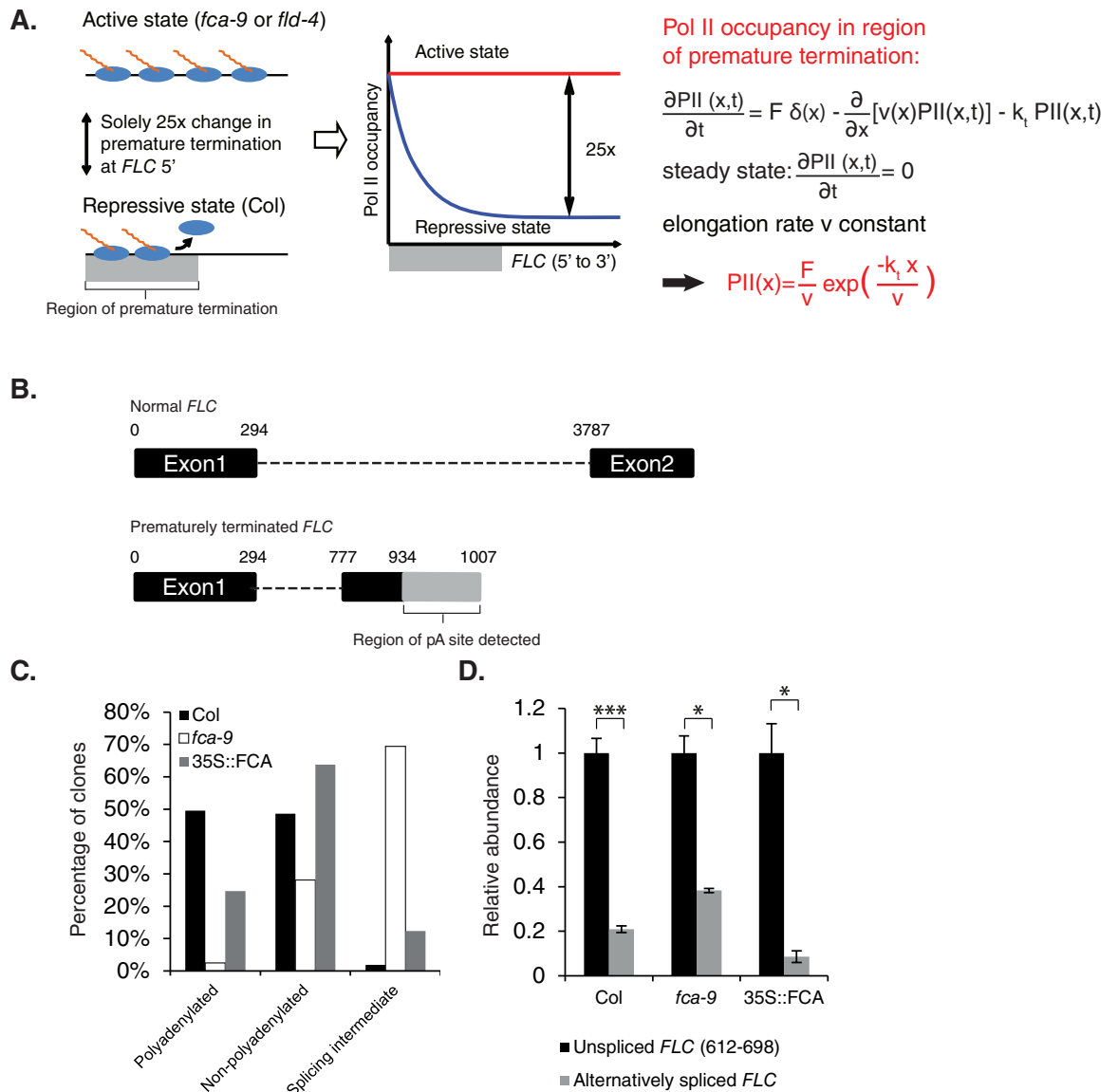


Figure 2.8: Detection of prematurely terminated *FLC* transcripts. (A) Schematic and mathematical analysis of influence of premature termination on sense Pol II occupancy across *FLC*. (B) Schematic of prematurely terminated transcripts detected from 3' RACE. Upper panel indicates the major *FLC* transcript with regions omitted after exon2. Lower panel indicates the splicing pattern of prematurely terminated *FLC* transcripts. Exons and introns are indicated by black boxes and dashed lines respectively. The grey box indicates a region containing polyadenylation sites detected in 3' RACE. Numbers indicate positions (bp post-TSS). (C) Percentage of polyadenylated, non-polyadenylated and splicing intermediate RNAs mapped in Col, *fca-9* and 35S::FCA. (D) Abundance of transcripts with the alternatively spliced intron in different genotypes. Levels were measured using a primer set with one primer spanning exon-exon junction. Values are mean±s.e.m. from 3 independent samples; data shown as abundance relative to intronic *FLC* level (612-698 bp post-TSS). For more details see [208]. Experiments performed by Zhe Wu.

more often in Col and 35S::FCA than in the *fca-9* mutant. This is however not the case (Figure 2.8D). Altogether, we conclude that sense premature termination occurs only occasionally at *FLC* and is not a major contributor to *FLC* repression.

### 2.2.8 *Co-transcriptional 5' to 3' degradation contributes little to FLC repression*

Next, we investigate theoretically a specific model based on 5' to 3' co-transcriptional RNA degradation [61] and initially focus on sense transcription only. We show that the *FLC* RNA fold changes (Figure 2.1) are also consistent with nascent RNA degradation rather than transcription initiation as a mode of regulation. Antisense transcription will be treated in subsection 2.4.3 in the context of model fitting to all above described experimental data. We will conclude that depending on the parameters, this model can be consistent with certain data sets in separation but, it is not consistent with all described data sets simultaneously. We therefore do not favour this model over the one with coordinated initiation and elongation rate changes, which is consistent with all available data and successfully predicted the intronic RNA fold change patterns.

The 5'-3' co-transcriptional degradation (coTD) model assumptions are as follows. Once Pol II has initiated transcription, the 5' end of the nascent, Pol II bound, RNA can be targeted by Xrn type 5'-3' exonucleases [61] during Pol II elongation with rate  $k_{tar}$  ( $s^{-1}$ ). For convenience we refer to this process also as targeting Pol II, but it does not affect the (assumed constant) Pol II elongation rate  $v$ . Normally nascent 5' RNA is capped soon after transcription initiation [47], thereby inhibiting the ability of Xrn to degrade nascent RNA. We do not model the (de)capping process but active decapping could be implicitly incorporated as part of the targeting rate  $k_{tar}$ . Once Pol II is targeted, Xrn will start degrading the nascent RNA from 5' to 3' with degradation rate  $k_d$  (units:  $bp/s$ ). Pol II keeps on elongating and Pol II terminates as soon as the Xrn has eaten up all the nascent RNA up to the then current Pol II position. The assumption that Pol II terminates at that moment is inspired by the established Xrn Torpedo mechanism of Pol II termination after mRNA cleavage at a

poly(A) site [47]. In case Pol II can elongate to the canonical 3' end of the gene before Xrn has reached it, we assume Pol II then pauses at the poly(A) site and terminates with the same rate  $k_{pA}$  as discussed previously. Note that this model is not falsified by the results described in subsection 2.2.7, because 5' to 3' RNA degradation would not be detected in the 3'RACE assay. For simplicity we assume that targeting cannot occur anymore once Pol II has transcribed intron1 so that splicing can occur. When we relax this assumption and allow targeting at any time before splicing occurs (results not shown), the conclusion that the model is inconsistent with the combined experimental data remains unaltered.

We will now derive steady state Pol II levels and RNA levels for this model. In this section we set TSS=0, so that all mentioned positions indicate distances from TSS, then the equation upstream of the intron1 acceptor site (I1A) for *untargeted* Pol II,  $PII_u(x)$ , is simply Equation 2.12 with  $k_t$  replaced by  $k_{tar}$ :

$$\frac{dPII_u(x)}{dx} = -\frac{k_{tar}}{v} PII_u(x).$$

Analogous to Equation 2.13, solving this equation analytically results in:

$$PII_u(x) = \begin{cases} \frac{F}{v} \exp\left(-\frac{k_{tar}}{v}x\right) & \text{if } 0 \leq x \leq I1A, \\ \frac{F}{v} \exp\left(-\frac{k_{tar}}{v}I1A\right) & \text{if } I1A \leq x < 3'. \end{cases}$$

Since targeted Pol II only drops off as soon as the Xrn have degraded all the corresponding nascent RNA, we define  $\Omega(x)$  as the Pol II position at the moment of Xrn targeting such that Pol II drops off at position  $x$ . So equating the time duration of Pol II elongation after targeting and Xrn RNA degradation we have  $\frac{x-\Omega(x)}{v} = \frac{x}{k_d}$  leading to  $\Omega(x) = x(1 - \frac{v}{k_d})$ . Note that we assume here  $k_d > v$ . See subsection 2.4.3 for discussion of the case  $k_d \leq v$ . The targeted Pol II distribution that is still at *FLC* at  $x$ ,  $PII_t(x)$ , then equates to the fraction of Pol II that was targeted between  $\Omega(x)$  and  $x$ :

$$PII_t(x) = \begin{cases} \frac{F}{v} \left[ \exp\left(-\frac{k_{tar}}{v}\Omega(x)\right) - \exp\left(-\frac{k_{tar}}{v}x\right) \right] & \text{if } 0 \leq x \leq I1A, \\ \frac{F}{v} \left[ \exp\left(-\frac{k_{tar}}{v}\min(\Omega(x), I1A)\right) - \exp\left(-\frac{k_{tar}}{v}I1A\right) \right] & \text{if } I1A \leq x < 3'. \end{cases}$$

By summing  $PII_u(x)$  and  $PII_t(x)$ , we then arrive at the total sense Pol II levels at steady state. We distinguish whether or not all targeted Pol

II have dropped off the gene template before reaching the 3' end:

$$\text{PII}(x) = \begin{cases} \frac{F}{v} \exp\left(-\frac{k_{tar}}{v} x \left(1 - \frac{v}{k_d}\right)\right) & \text{if } 0 \leq x < \min\left(\frac{IIA}{1 - \frac{v}{k_d}}, 3'\right) \\ \frac{F}{v} \exp\left(-\frac{k_{tar}}{v} IIA\right) & \text{else.} \end{cases} \quad (2.14)$$

Now we assume for simplicity that once the Pol II gets to the 3' end, Pol II can only drop off through regular poly(A) processing and not through nascent RNA degradation-induced torpedo mechanism. However it must be noted that the eventual conclusion that this model cannot fit the data, will not depend on this assumption (see subsection 2.4.3 for more details). Analogous to Equation 2.5, the Pol II levels at the 3' poly(A) site then result in:

$$\text{PII}(x) = \begin{cases} \frac{F}{L_{PII} k_{pA}} \exp\left(-\frac{k_{tar}}{v} 3' \left(1 - \frac{v}{k_d}\right)\right) & \text{if } 3' < \frac{IIA}{1 - \frac{v}{k_d}} \\ \frac{F}{L_{PII} k_{pA}} \exp\left(-\frac{k_{tar}}{v} IIA\right) & \text{if } \frac{IIA}{1 - \frac{v}{k_d}} \leq 3'. \end{cases}$$

In this coTD model, the RNA production rate  $p$  is determined by the flow of untargeted Pol II reaching the intron1 acceptor site:  $p = F \exp\left(-\frac{k_{tar}}{v} IIA\right)$ .

We now proceed with determining the intronic RNA levels at position  $z$  arising from transcription. We differentiate between RNA bound by untargeted Pol II located in intron1,  $RNA_u(z)$ ; RNA with the complete intron transcribed that is competent to be spliced out,  $RNA_s$ ; RNA that is in the process of being degraded bound to targeted Pol II  $RNA_t(z)$  and lariat intron RNA  $RNA_{lariat}(z)$  that is in the process of being degraded. Similar to the equations in section subsection 2.2.5 we have

$$RNA_u(z) = \int_z^{IIA} \text{PII}_u(y) dy = \frac{F}{k_{tar}} \left[ \exp\left(-\frac{k_{tar}}{v} z\right) - \exp\left(-\frac{k_{tar}}{v} IIA\right) \right].$$

Furthermore  $RNA_s = \frac{F \exp\left(-\frac{k_{tar}}{v} IIA\right)}{k_s}$ . The 5' RNA of the targeted Pol II is in the process of degradation, so what RNA is still present depends on when the Pol II was targeted. We define  $\Theta(z, x)$  as the targeting position such that RNA is degraded up to  $z$  while the Pol II is currently at  $x \geq z$ . By equating time scales we obtain  $\frac{x}{v} = \frac{z}{k_d} + \frac{\Theta(z, x)}{v}$ , and thus  $\Theta(z, x) = x - z \frac{v}{k_d}$ . When we take into account that targeting can only occur up to  $IIA$ , then the density of targeted Pol II at position  $x$  that still has RNA corresponding to  $z$  attached equals  $\frac{F}{v}$  times the



fraction of Pol II that was targeted between  $\min[\Theta(z, x), l1A]$  and  $\min[x, l1A]$ :

$$\text{PII}_t^z(x) = \frac{F}{v} \left[ \exp\left(-\frac{k_{tar}}{v} \min[\Theta(z, x), l1A]\right) - \exp\left(-\frac{k_{tar}}{v} \min[x, l1A]\right) \right].$$

Now we can find the associated targeted RNA fraction that still contains RNA corresponding to position  $z$  by integrating. We note that beyond  $\frac{l1A}{1-\frac{v}{k_d}}$ , all targeted Pol II has terminated already, so that  $\text{PII}_t^z(x) = 0$  for  $x > \frac{l1A}{1-\frac{v}{k_d}}$ . For certain parameters it could be that the position  $\frac{l1A}{1-\frac{v}{k_d}}$  extends beyond the sense termination site  $\left(\frac{l1A}{1-\frac{v}{k_d}} > 3'\right)$ , so that Pol II might terminate due to regular 3' termination rather than termination due to nascent RNA degradation. However, the precise Pol II termination mechanism does not affect the RNA contribution that we calculate here. Therefore we find:

$$\begin{aligned} \text{RNA}_t(z) &= \int_z^{\frac{l1A}{1-\frac{v}{k_d}}} \text{PII}_t^z(y) dy \\ &= \frac{F}{k_{tar}} \left[ \exp\left(-\frac{k_{tar}}{v} z \left(1 - \frac{v}{k_d}\right)\right) - \exp\left(-\frac{k_{tar}}{v} z\right) \right] \\ &\quad - F \exp\left(-\frac{k_{tar}}{v} l1A\right) \frac{z}{k_d}. \end{aligned}$$

Analogous to our results in subsection 2.2.5, the lariat RNA is formed by  $\text{RNA}_{lariat}(z) = F \exp\left(-\frac{k_{tar}}{v} l1A\right) \frac{z-l1D}{k_l}$ . In conclusion we have for the total intronic RNA in a coTD model:

$$\begin{aligned} \text{RNA}_{total}(z) &= \text{RNA}_u(z) + \text{RNA}_t(z) + \text{RNA}_s + \text{RNA}_{lariat}(z) \\ &= \frac{F}{k_{tar}} \left[ \exp\left(-\frac{k_{tar}}{v} z \left(1 - \frac{v}{k_d}\right)\right) - \exp\left(-\frac{k_{tar}}{v} l1A\right) \right] \\ &\quad + F \exp\left(-\frac{k_{tar}}{v} l1A\right) \left[ \frac{-z}{k_d} + \frac{1}{k_s} + \frac{z-l1D}{k_l} \right]. \end{aligned}$$

As discussed in the previous sections, the nascent RNA fraction is the total minus the lariat RNA. For nascent exon1 RNA we retain the expressions for  $\text{RNA}_u(z)$  and  $\text{RNA}_t(z)$ , and replace  $k_s$  by  $k_x$ .

A detailed parameter fitting procedure described in subsection 2.4.3 indicates that this coTD is not quantitatively consistent with all available experimental data. Intuitively, this results from the requirement of a sufficiently high  $k_{tar}$  and  $k_d$  so that most Pol II in Col terminates prematurely, which causes a significant drop in Pol II levels. This is inconsistent with the observed Pol II fold changes. Interestingly

repeating the fitting procedure for a 'hybrid' model whereby we allowed for coTD and initiation rate changes ( $\frac{F^{fc-9}}{F^{Col}} = 2$  or 5), resulted in the same conclusion. This is because the elongation rate changes are determined by the nascent RNA fitting procedure and appear inconsistent with the Pol II fold changes.

In a coTD model we would expect that absence of degradation ma-

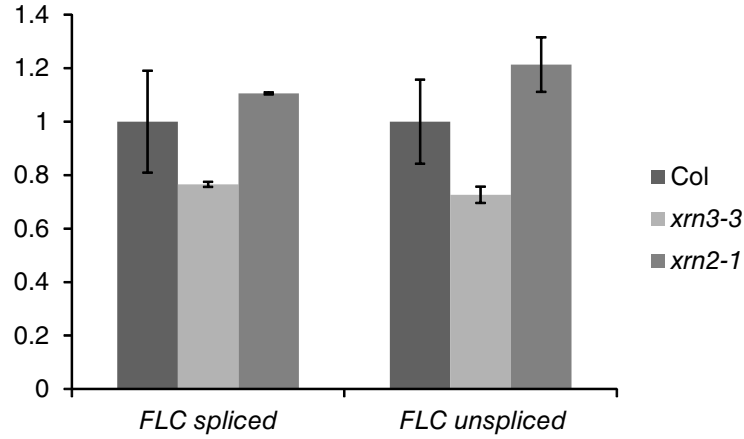


Figure 2.9: Relative RNA expression level of *FLC* spliced and unspliced in Col, *xrn3-3* and *xrn2-1*. Level in Col was set as one. Values are mean  $\pm$  standard deviation from 3 independent samples. For more details see [208]. Experiments performed by Zhe Wu.

chinery would lead to de-repression of *FLC* expression. However, in *xrn2-1* (null) and *xrn3-3* (hypomorphic) mutants [217], no significant increase in *FLC* unspliced and spliced RNA is observed (Figure 2.9, for more details see [208]). We also investigated mathematically 3'-5' degradation after endonucleolytic cleavage of Pol II-bound RNA. However, besides the experimental Pol II upregulation being inconsistent, an extensive experimental mapping of RNA 3' ends (subsection 2.2.7) did not support this mechanism playing a major role in *FLC* regulation. RNA degradation at the locus after elongation is also quantitatively inconsistent with the experimental data for the same reasons as the nascent 5'-3' RNA degradation model. We conclude that a coTD model and other mechanisms based on RNA degradation are not consistent with the available experimental data.

Instead we favour the model with coordinated initiation and elongation changes.

### 2.2.9 *FLC transcriptional repression involves local FLD activity and an altered chromatin state*

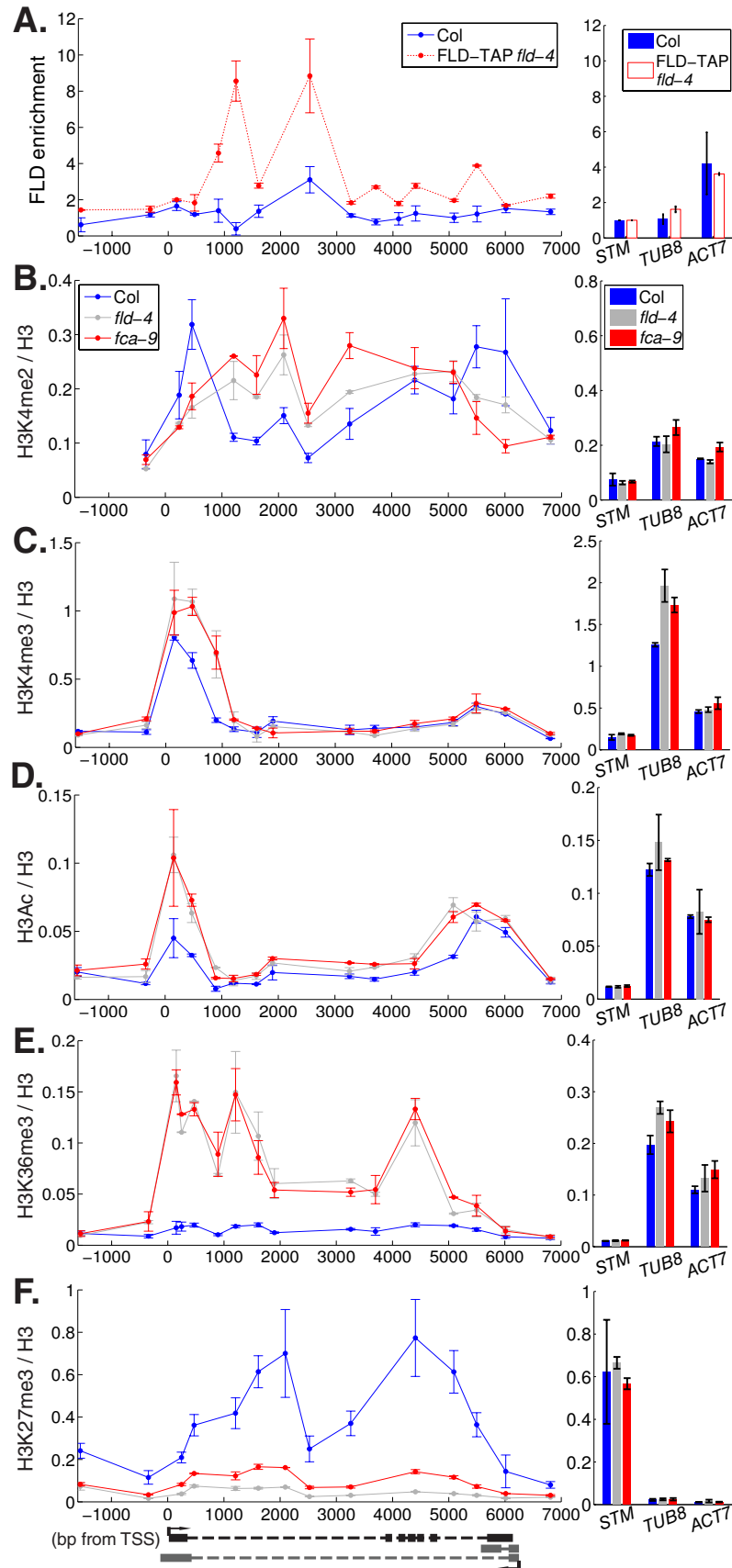
We next investigated how the coordinated change in initiation and elongation rates between the active and repressed states at *FLC* is connected with the chromatin state [208], bearing in mind that FLD is a histone demethylase that is genetically epistatic to FCA [120]. We found that FLD is targeted to the entire *FLC* gene with highest enrichment around 1kb to 3kb downstream of the TSS (Figure 2.10A). This localization is consistent with the increased H3K4me2 in the *FLC* gene body (1kb to 4kb beyond the TSS) in the *fld-4* mutant (Figure 2.10B). This 2-fold increase of H3K4me2 was paralleled by changes in other histone modifications. In both *fld-4* and *fca-9*, H3K4me3 and H3Ac were increased mostly around the *FLC* sense TSS, coincident with lower H3K4me2 in this region (Figure 2.10B-F). H3K36me3 increased along the whole gene (Figure 2.10E), with an opposing pattern to changes in H3K27me3 (Figure 2.10F). For more details see [208].

### 2.2.10 *Stochastic model based on FLD action through FCA-dependent proximal COOLAIR processing*

Up to this point, our modelling has not attempted to address how FLD and FCA mechanistically achieve repression in Col. For instance it remains unclear how proximal polyadenylation of *COOLAIR* relates to *FLC* gene body histone methylation via FLD. More generally, how the overall chromatin state affects sense *FLC* expression is unknown. In this section we detail how a *COOLAIR* -mediated chromatin mechanism could coordinate transcriptional initiation and elongation to repress *FLC* expression.

From a computational perspective, we implemented a Gillespie algorithm [218] in C++. The reactions included are described below and illustrated in Figure 2.11. We show this model is consistent with the available data and predict that FCA-dependent *COOLAIR* proximal termination, the local chromatin state and slow Pol II elongation

Figure 2.10: FLD enrichment at the *FLC* locus is associated with changed histone modifications. (A) FLD-TAP ChIP enrichment across *FLC* in Col and FLD-TAP/*fld-4*. Values are mean  $\pm$  s.e.m. from 2 independent samples, with data presented as enrichment at *FLC* relative to enrichment at control gene *STM*. (B-F) ChIP across *FLC* in Col, *fca-9* and *fld-4* measuring H3K4me2 (B), H3K4me3 (C), H3Ac (D), H3K36me3 (E), H3K27me3 (F). Values are mean  $\pm$  s.e.m. from 2 independent samples, with data normalized to H3. Values at the control genes *STM*, *ACT7* and *TUB8* are shown on the right. For more details see [208]. Experiments performed by Zhe Wu.



are all required and reinforce each other to achieve stable *FLC* repression.

Based on the available genetic and biochemical evidence we assume in the model that proximal antisense polyadenylation in the presence of FCA leads to recruitment of FLD to the locus. FLD, a histone H3K4me2 demethylase [120], then represses levels of an 'activating' histone modification, referred to as the A mark [113, 138]. This A mark stimulates sense and antisense transcriptional initiation and elongation. Thus the A mark is positively correlated with the transcriptional activity at the *FLC* locus. Note that the exact molecular nature of the A mark is not critical for the repression mechanism, but could correspond, for example, to H3K36me3 levels (see subsection 2.2.9). We specifically track the number of FLD and A mark levels present at the locus. The A mark can assume values between  $H = 0$  and  $H = 70$ , approximately the number of H3 histones at the *FLC* locus [138].

The *FLC* locus discretization is as described in subsection 2.2.4 (see also Figure 2.11 and Table 2.1 for parameter values). Each site can be occupied by at most one Pol II. If a TSS site is unoccupied, a Pol II from the nucleoplasmic, free Pol II population, which numbers  $\text{PII}_{\text{free}}$ , can bind to that TSS, resembling transcription initiation. Once a Pol II has bound to a sense (antisense) TSS we assume it is competent to elongate in the sense (antisense) direction. If the TSS site is unoccupied, sense initiation can occur with propensity  $p_t$ , that depends linearly on the A mark level and on  $\text{PII}_{\text{free}}$  (with reaction rate  $k_{f,s}$ ) above a threshold  $T_{f,s}$ :

$$p_t = \begin{cases} k_{f,s} \text{PII}_{\text{free}} (A - T_{f,s}) & \text{if } A > T_{f,s} \\ 0 & \text{otherwise} \end{cases}$$

For antisense initiation, we used the same function for the propensity, but with threshold  $T_{f,as}$  and reaction rate  $k_{f,as}$  instead.

As for the analytical coordination model, we assume the magnitude of the elongation rate is independent of both direction (sense or antisense) and position along the *FLC* gene. Importantly, we do assume that the magnitude of the elongation rate is dependent on the

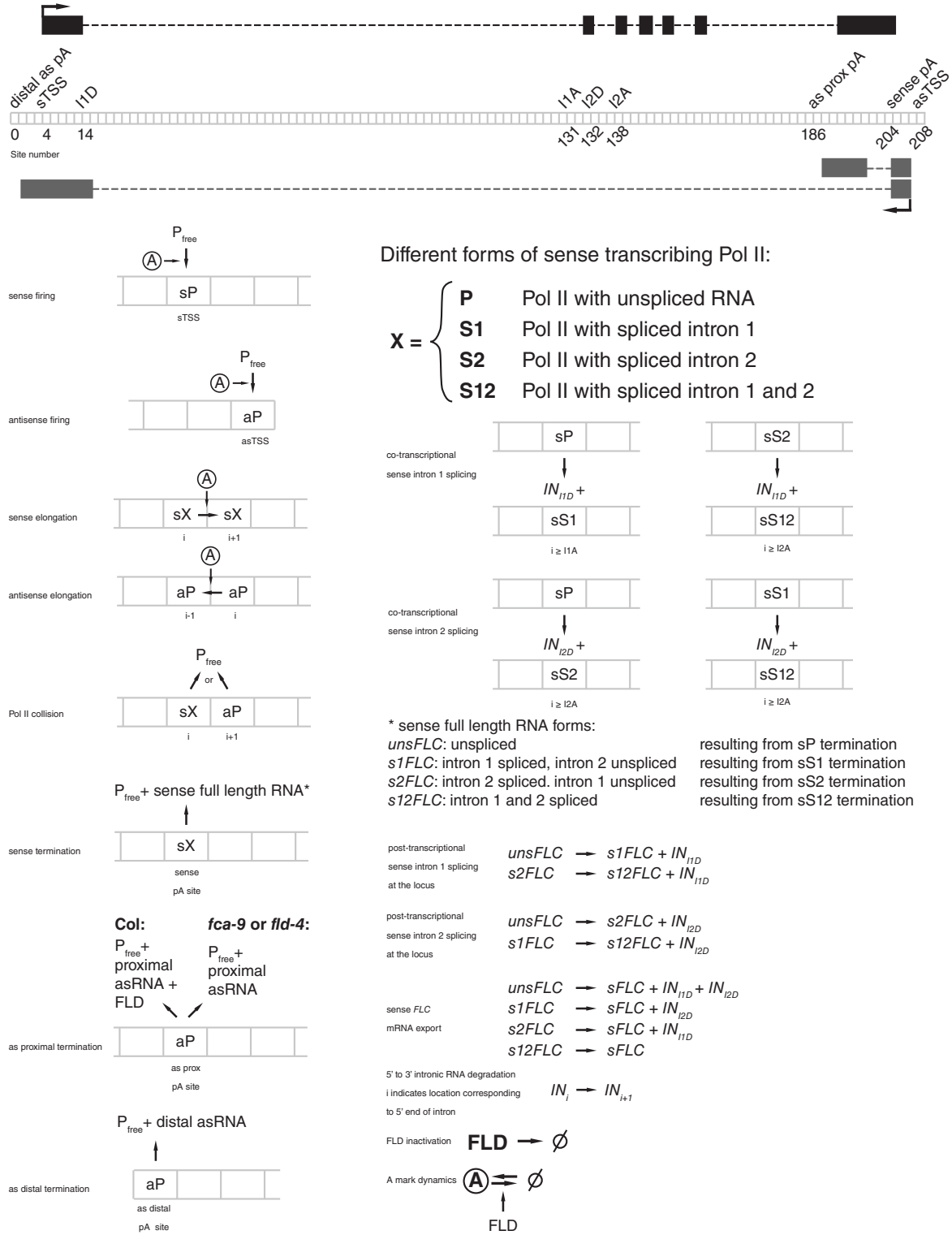


Figure 2.11: Schematic of the processes taken into account in the stochastic simulations of *FLC* transcriptional repression, with coordinated initiation and elongation rates. Black (grey) rectangular boxes in the schematic indicate sense (antisense) exons.

A mark level (see below for the biological interpretation of this assumption). We use a similar functional form as for the initiation propensity above. In case of sense (antisense) transcription, a Pol II at site  $i$  ( $P_i = 1$ ) can elongate to a neighbouring site  $i + 1$  ( $i - 1$ ), if that neighbouring site is unoccupied, with propensity:

$$p_t = \begin{cases} k_{\text{elon}} P_i (A - T_{\text{elon}}) & \text{if } A > T_{\text{elon}} \\ 0 & \text{otherwise} \end{cases}$$

Again, we assume that termination of a transcribing Pol II can only occur at the aforementioned sense, and proximal and distal antisense termination sites (Figure 2.11). When a sense Pol II reaches the sTES, it can terminate with probability per unit time  $k_{pA}$ , resulting in a free Pol II and a 3' processed sense transcript that remains at the locus (Figure 2.11). For more details on the different reactions involving sense RNA, see below. Similarly, antisense elongation cannot proceed beyond the distal poly(A) site (Figure 2.11). When an antisense Pol II reaches the distal poly(A) site, also with a probability per unit time  $k_{pA}$ , it can terminate, resulting in a free Pol II and the production of a processed distal antisense RNA.

FCA localizes around the proximal antisense poly(A) site at the *FLC* locus [120], likely binding to the antisense RNA. FCA interacts with 3' processing factors to facilitate proximal polyadenylation [98, 115] and splicing [127]. We model FCA's action implicitly to act in concert with 3' processing factors as follows (Figure 2.11): when an antisense Pol II reaches the proximal termination site, it can continue to elongate or, whilst residing at this site during a 'window of opportunity' [44], it can terminate with a probability per unit time  $k_{pA}$ . The latter process results in a free Pol II, production of a processed proximal antisense RNA and the recruitment of one FLD molecule to the locus. Note that the elongation rate affects the time Pol II spends inside this window and therefore the probability that it produces a proximal transcript and recruits FLD. In an *fca-9* mutant we assume termination and proximal processing can still occur with the same probability per unit time  $k_{pA}$  at the same site. However, we assumed in an *fca-9* mutant background that this reaction does not lead to FLD recruitment. So in the model, FLD recruitment is FCA-dependent. It remains unclear whether this assumption is actu-

ally true. However, the mechanism described here can also be interpreted as FLD *activity* at the locus being FCA-dependent: if the FLD recruitment step in the model is interpreted as inducing the activity of FLD at the locus. An FLD ChIP experiment in an *fca-9* mutant would not be able to falsify this latter interpretation. We continue the model description using the terminology of FLD recruitment to and removal from the locus.

FLD disappearance from the locus is assumed to occur in a first order reaction with a probability per unit time of  $k_{FLD}$  for each FLD molecule. 'A' mark placement occurs constitutively with probability per unit time  $p_t = k_A(H - A)$  depending on the available number of free histone tails and the parameter  $k_A$ . 'A' mark removal reactions occur with a probability per unit time  $p_t = (k_{r,0} + k_{r,s}FLD)A$ , depending on the number of 'A' marks, a parameter  $k_{r,0}$ , reflecting the histone turnover time, as well as an FLD-stimulated removal term  $k_{r,s}FLD$ , where  $FLD$  is the number of FLD proteins at the locus.

In case of polymerase collision, with a sense Pol II at site  $i$  and an antisense Pol II at  $i + 1$ , we assume that with probability per unit time  $k_{col}$  the sense Pol II can be evicted off the gene resulting in a free Pol II. Likewise, with the same probability per unit time, the antisense Pol II can be evicted. We note, however, that transcriptional interference cannot be the main mechanism for FCA pathway regulation, as it is incompatible with the positive correlation between sense and antisense RNA transcripts.

The production of sense RNA is modelled as follows (Figure 2.11). A sense Pol II at site  $i$  has produced unspliced RNA corresponding to the sites *TSS* up to  $i - 1$ . Splicing of sense *FLC* intron1 and intron2 are explicitly modelled: as soon as Pol II elongates past the intron1 acceptor site *I1A*, Pol II can continue to elongate and, in addition, splicing of intron1 can occur with a probability per unit time  $k_s$ . Note that this parameter is the only one allowed to differ between *fca-9* and *fld-4*. In the latter case, RNA corresponding to sites from the intron1 donor site *I1D* up to *I1A* - 1 are now cleaved intronic RNA with the 5' end at site *I1D*. We term this RNA species  $IN_{I1D}$ .



with the index indicating the 5' end of the intronic RNA. This RNA can then be degraded from 5' to 3' (Hesselberth, 2013) with reaction rate  $k_l$  [44], where the 5' to 3' degradation of intronic RNA corresponding to site  $j$  occurs with propensity  $p_t = k_l IN_j$ . As a result of this reaction,  $IN_{j+1}$  is formed, corresponding to intronic RNA with a 5' end at site  $j + 1$  (and 3' end at  $I1A$ ). Degradation of  $IN_{I1A-1}$  occurs without a reaction product, representing the last step in the 5 to 3' degradation of an intronic RNA. Similarly, intron2 splicing can occur with probability per unit time  $k_s$  as soon as a Pol II reaches site  $I2A = 138$ , resulting in  $IN_{I2D}$ , a cleaved intron 2 RNA with a 5' end at the intron 2 donor site  $I2D = 132$  and a 3' end at  $I2A$ . Degradation of this cleaved intron2 RNA occurs in the same way as with intron1 degradation. We only explicitly modelled splicing reactions of sense intron1 and 2 in the simulations, which are assumed to be independent reactions. Incorporating splicing of additional sense introns in the model would not affect our results provided that each splicing reaction is independent of the others.

Splicing can also occur after sense Pol II has terminated (Figure 2.11). If Pol II has terminated with both introns spliced out,  $s12FLC$  is created, a cleaved full length RNA at the locus with both intron1 and 2 spliced out. If Pol II has terminated prior to splicing of intron1 and 2, a full length, unspliced RNA termed  $unsFLC$  is produced.  $unsFLC$  can be spliced in intron1 or 2 with probability per unit time  $k_s$  with the same intronic RNA degradation process as for Pol II-bound RNA. If Pol II terminates with only intron1 (2) spliced out, then  $s1FLC$  ( $s2FLC$ ), is produced, indicating a full length RNA with only intron1 (2) spliced out. Splicing of intron2 (1) can then occur, again with probability per unit time  $k_s$ . All full length transcripts,  $unsFLC$ ,  $s1FLC$ ,  $s2FLC$  and  $s12FLC$  can be exported with probability per unit time  $k_x$  leading to a mature (spliced) sense  $FLC$  mRNA:  $sFLC$ . As part of these reactions, any as yet unspliced intron1 and/or 2 is then also assumed to be spliced out, resulting in the intron degradation products  $IN_{I1D}$  and/or  $IN_{I2D}$ . However, since the timescale of splicing is much shorter than the combined timescale to elongate across the locus and then be exported (Table 2.1), this assumption does not significantly affect our results. In addition, splicing dynam-

ics for the antisense transcripts are not included in our simulations, as we do not fit any unspliced antisense RNA data.

Simulations started at time  $t = 0$  and ran until (simulated) time  $t$ , updated according to the Gillespie algorithm, exceeded a predefined time of 10 days. To let the system reach steady state, we started to output simulation data (RNA, Pol II and A mark levels) after  $t = 5$  days at regular time intervals of 100s. From the *sFLC* levels between  $t = 5$  days and  $t = 10$  days we inferred the (time-averaged) *FLC* mRNA production rate. The ratio of production rates of *fca-9* or *fld-4* over Col could then be compared to the experimental spliced *FLC* fold changes from the total RNA assays. Similarly we could compare the proximal and distal transcript fold changes with the experimental spliced antisense forms (Figure 2.12A). Note that this procedure involving ratios did not need to assume any particular value for the mRNA degradation rates, only that these rates remain unchanged between Col and *fca-9/fld-4* (as discussed in previous sections).

The simulated Pol II levels (arising from summing the sense and antisense Pol II) were also summed over all time points between  $t = 5$  days and  $t = 10$  days and then convoluted as described in subsection 2.4.2 for direct comparison with our experiments (Figure 2.12B,C). In addition, we used the output time course data on A levels and Pol II occupancy along the gene, in combination with the above described formulas for the initiation and elongation propensities, to generate the distribution of propensities for these processes in the presence of potential Pol II occlusion/collision effects and A levels that are specific to Col and *fca-9/fld-4*. From these propensity distributions we subsequently calculated the respective mean values, which are estimates of the mean initiation and elongation rates. We also directly kept track of the total number of (sense and antisense) initiation events that occurred in the simulations between  $t = 5$  days and  $t = 10$  days. Extracting the actual mean elongation rates from the simulations directly is technically difficult as that would require keeping track of the duration and distance between initiation and drop off for each Pol II. However, the estimated mean initiation rates coincide within 0.1% of the actual mean initiation rates. This sug-

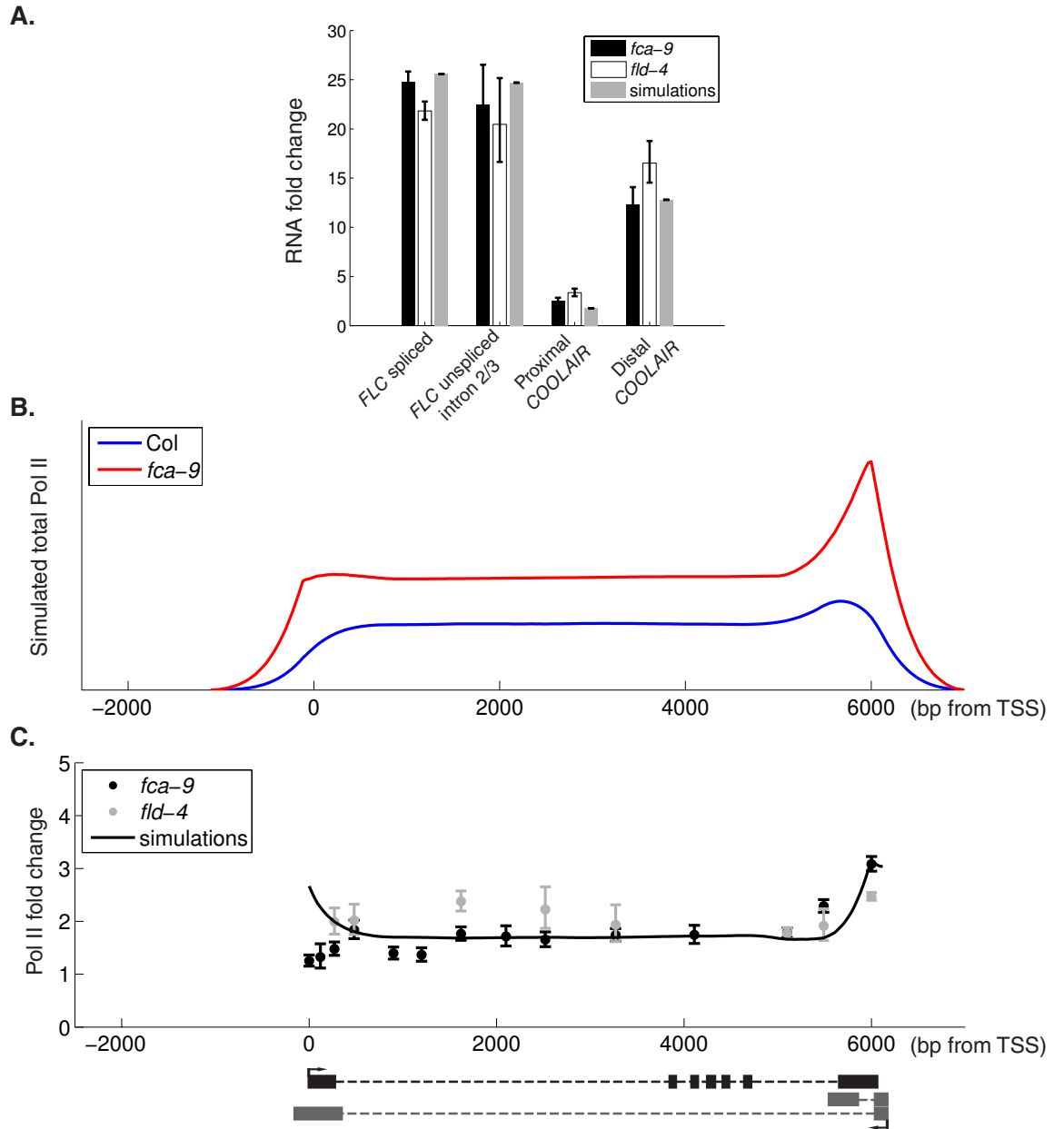


Figure 2.12: Stochastic coordination model can fit the prior RNA and Pol II data. (A) RNA fold up-regulation in *fca-9* and *fld-4* mutants compared to Col: experimental data as in Figure 2.1. The model values are the fits to the experimental data. (B) Total Pol II levels in Col and *fca-9* across *FLC* from the model convolved with experimental Pol II ChIP fragment size distribution. (C) Experimental and model Pol II fold upregulation. Experimental data as in Figure 2.4D. Model fold changes are ratio of profiles shown in (B).

gests that the mean rates in Col and the *fca-9/fld-4* mutant estimated from the time course data accurately reflect the actual mean rates resulting from the stochastic simulation outcomes. The time course of simulated A mark levels also generates a distribution of values. We therefore calculated the ratio of simulated mean A mark levels between Col and the *fca-9* mutant to compare the simulated fold upregulation with the experimental mean H3K36me3 fold upregulation across the *FLC* locus ( $\sim 5x$ , see Figure 2.10E), yielding good agreement with the model.

The simulated intronic RNA levels (arising from summing the relevant Pol II bound RNA, *unsFLC*, *s1FLC* and *s2FLC* and cleaved intronic RNA) for each site were averaged over all time points between  $t = 5$  days and  $t = 10$  days. Again the ratio of this quantity between *fca-9* or *fld-4* and Col can be compared with the experimental intronic RNA data from the total RNA assay Figure 2.13. To compare the simulations with the experimental chromatin-bound RNA assay, we omitted in the intronic RNA summation the cleaved intronic RNA [212]. Note that in these two cases, due to changes in  $k_s$ , our simulations differ between *fca-9* and *fld-4*, in contrast to all other situations where changes in  $k_s$  have no effect. In the chromatin-bound case we also computed simulated chromatin-bound *FLC* exon1 RNA levels to compare with the available experimental data. To obtain precise estimates of the simulated levels and fold upregulation, we averaged the procedures described above over  $10^3$  simulations.

Given the number of model parameters and the duration of simulations, an automated procedure to fit our stochastic simulations to the data was not practical. Instead we manually fitted the simulations to the quantitative RNA (Figure 2.1) and Pol II fold changes (Figure 2.4). Unless otherwise stated the *fld-4* parameters and results are the same as those for *fca-9*. Fitting the RNA fold changes (Figure 2.12) determines  $T_{f,s}, k_{f,s}, T_{f,as}, k_{f,as}$  only up to the mean relative initiation rates  $\frac{F^{fca-9}}{F^{Col}}$  (sense and antisense specific), but not absolute (s/as specific) initiation rates  $F^{Col}$  and  $F^{fca-9}$ . Although Pol II ChIP is not a strand specific technique, the quantitative Pol II fold upregulation (Figure 2.12) near the sense and antisense 3' end could

only be fitted if  $F_s^{fca-9} \approx F_{as}^{fca-9}$ . The absolute initiation rates were not, however, constrained by this procedure. We chose  $T_{f,s}, k_{f,s}, T_{f,as}, k_{f,as}$  such that the initiation rates were as large as possible (to minimise FLD dwell time at *FLC*), but without transcriptional interference affecting the regulation of *FLC*. Higher absolute initiation rates would lead to the following situation: due to transcriptional interference mostly in *fca-9*, the fold change in initiation rate  $\frac{F_{fca-9}}{F_{Col}}$  would be then be higher than the RNA production rate fold change (as measured by how many *sFLC* is produced) which turns out to be inconsistent with our Pol II data. This procedure resulted in mean rates:  $F_s^{Col} = 0.01 min^{-1}$ ,  $F_s^{fca-9} = 0.25 min^{-1}$ ,  $F_{as}^{Col} = 0.012 min^{-1}$ ,  $F_{as}^{fca-9} = 0.17 min^{-1}$ . The elongation rate values are also not determined in absolute terms by Pol II ChIP, as this technique only measures relative levels. The absolute elongation rate values scale directly with the termination rate  $k_{pA}$ , in order to fit the quantitative fold changes of Pol II both in the body and at the 3' end of *FLC* (Figure 2.12C). In the absence of further information about the termination rate at *FLC*, we chose  $k_{pA} = \frac{1}{50} s^{-1}$  [41]. Note that we do not need a different  $k_{pA}$  value for the distal antisense in order to fit the model to the data, unlike in the analytical model (see subsection 2.2.4). This difference arises because in the more detailed, stochastic model transcriptional interference and different mean elongation rates between sense and antisense can occur (see below).

Then by fitting the quantitative Pol II changes in the gene body (1.8x) we determined  $T_{elon}, k_{elon}$ , resulting in  $v^{Col} = 0.13 kb/min$  and  $v^{fca-9} = 1.5 kb/min$  and thus a 12x mean elongation rate change. In our simulations, the sense specific Col elongation rate  $v_s^{Col} = 0.23 kb/min$  is somewhat higher than  $v^{Col}$  arising from sense and antisense elongation. This difference arises due to the coordination between initiation and elongation: in Col, sense initiation (and subsequent elongation) only occurs when the A mark level is (transiently) above  $T_{f,s}$ , the initiation threshold that itself is considerably higher than  $T_{f,as}$  (Table 2.1). Thus sense elongation coincides with periods of higher levels of A and thus quicker elongation. Similarly due to coordination between initiation and elongation, antisense initiation occurs even at very low levels of A resulting in the antisense specific elon-

gation rate  $v_{as}^{Col} = 0.09 \text{ kb/min}$ . With sense and antisense elongation in *fca-9* being equal to the mean elongation rate ( $v_s^{fca-9} = v_{as}^{fca-9} = 1.5 \text{ kb/min}$ ), our predicted sense (antisense) specific elongation rate changes are thus 7x (17x). We conclude that in order to fit the simulations to the prior RNA and Pol II data, we require elongation rate changes of 7 – 17 fold.

Turning to the experimental chromatin-bound RNA fold changes across *FLC* (Figure 2.13), we found that no additional parameter tuning was needed to explain this data since the parameter  $k_s = 0.01 \text{ s}^{-1}$  was estimated from the literature [44, 213], the initiation rate upregulation  $\frac{F_s^{fca-9}}{F_s^{Col}}$  was already determined by the spliced *FLC* upregulation, and the elongation rate estimates were already obtained by fitting stochastic simulations to the Pol II ChIP assay. Furthermore, for total RNA fold changes across *FLC* (Figure 2.13), we found that a value of  $1.5 \text{ bp/s}$  for the 5' to 3' lariat degradation rate  $k_l$  is consistent with reported intron half-lives in mammals [44]. This value could then predict our total RNA levels in addition to our other data. In *fld-4* all parameter values were as in *fca-9*, with the exception of the splicing parameter which was increased two fold as in the analytical model.

To verify that our stochastic and analytical approaches generated similar results (Figure 2.13), we took the following approach: insertion of our mean sense initiation ( $F_s^{Col}, F_s^{fca-9}$ ) and sense elongation rate ( $v_s^{Col}, v_s^{fca-9}$ ) estimates arising from the stochastic simulations, as well as values for  $k_l$  and  $k_s$ , directly into the analytic equations Equation 2.8 and Equation 2.9 for RNA fold changes across *FLC* intron1, generated profiles that are very similar to those resulting from our simulations (Figure 2.13). Hence, the stochastic simulations and the analytical derivation both correctly predicted the experimental *FLC* intron1 upregulation, thereby validating elongation rate changes of at least 7x. This result also indicates that, given the mean initiation and elongation rates, stochastic fluctuations do not systematically alter the mean RNA fold upregulation, which results from averaging over a sufficiently large time window and number of simulation outcomes.

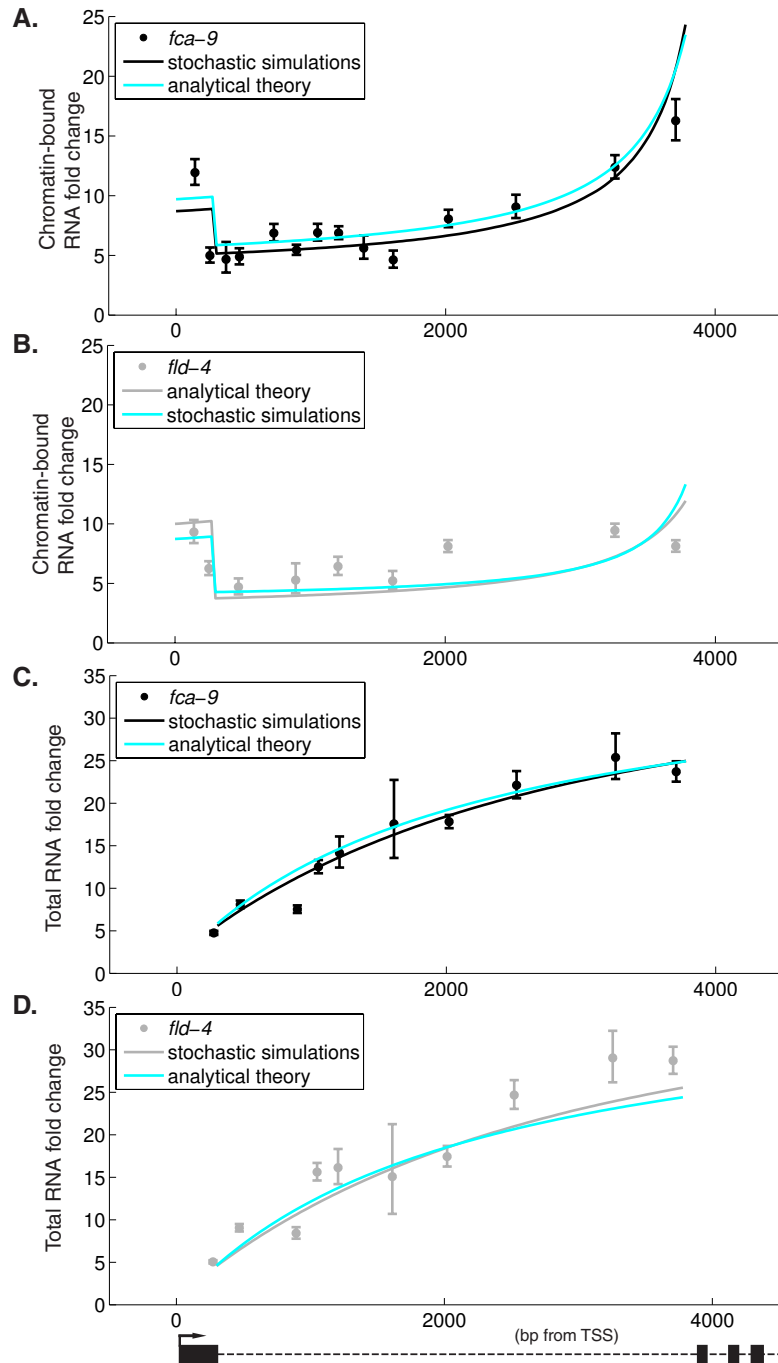


Figure 2.13: Stochastic simulations and analytical theory generate similar results. Graphs showing predicted *FLC* intron1 RNA fold change (chromatin-bound and total RNA) according to stochastic simulations and analytical model, as well as the experimentally observed profiles (as in Figure 2.6). For the chromatin-bound RNA fraction, exon1 fold upregulation is also shown.

It has been proposed that the act of transcription to the distal site could alter the *FLC* chromatin state resulting in transcriptional activation [97]. We have modelled a scenario where FLD recruitment to the locus is constitutive and removed by sense and antisense transcription elongation through the gene body (results not shown). However this mechanism appeared unable to generate the relatively large RNA fold changes between Col and the *fca-9* or *fld-4* mutant (Figure 2.1). We therefore do not favour the hypothesis that proximal termination merely functions to prevent transcriptional readthrough which would then allow for chromatin silencing. We conclude that the stochastic coordination model based on FLD recruitment through FCA-dependent proximal *COOLAIR* processing is consistent with all the above described experimental data.

#### 2.2.1.1 *Stochastic coordination model predicts that slow Pol II elongation is required for FLC repression*

An important component of the above detailed *FLC* repression mechanism is the involvement of *COOLAIR* antisense transcripts and the Pol II elongation rate. As detailed in subsection 2.2.2, the elongation generally does not affect the RNA production rate. However, in our simulations (Figure 2.12A), slow elongation resulting from an FLD-induced repressive chromatin state is critical to explain *COOLAIR* and sense *FLC* expression. Slow elongation ensures that Pol II reaches the distal region more slowly, favoring splicing at the relatively weak acceptor site of the small antisense intron (Figure 2.14A) [127]. Splicing can then couple to proximal polyadenylation/termination of antisense transcription. In this way, both FCA recruitment and a repressive chromatin state are required to favor proximal 3' processing. Such proximal processing is associated with FLD recruitment (or FLD activation), reinforcing a repressive chromatin environment with reduced sense and antisense initiation and slow Pol II elongation, with the latter favoring proximal processing in a feedback loop (Figure 2.14B). The kymographs (Figure 2.14A) with the *FLC* gene on the horizontal axis and time running from top to bottom show how sense and antisense Pol II elongate through the gene over time in both Col and the *fca-9* mutant. To verify that slow Pol II elonga-



tion is required for repression in the stochastic model, we imposed a constant fast elongation rate independent of the chromatin environment, with parameters otherwise unchanged from our Col simulations (Figure 2.14A). As a result, we found that *FLC* expression is upregulated  $\sim 6\times$  compared to Col, indicating that without FLD-mediated feedback from the chromatin environment to elongation, repression would be compromised.

### 2.3 Discussion

Understanding how flowering time in plants is regulated has led into a detailed mechanistic dissection of the *A. thaliana* floral repressor *FLC*. Genetic screens have identified RNA processing factors of anti-sense transcripts to *FLC*, termed *COOLAIR*, and a histone demethylase FLD as central to the *FLC* repression mechanism. Here, using a combination of mathematical modeling and experiments, we have shown that this repression of *FLC* involves coordination of transcriptional initiation with elongation associated with a changed chromatin state. This conclusion is supported by genome-wide correlations found in yeast and mammalian cells between gene expression, gene body Pol II levels and Pol II elongation rates [17, 205].

The results from this study progresses our understanding of mechanisms important for developmental timing in plants. They also raise further questions: is there generally a need to coordinate initiation and elongation and how is this achieved? Control of gene expression may necessitate such coordination. For instance, 5' paused Pol II limits the number of additional Pol II molecules that can initiate [221]. However, this is not a feature at *FLC*, as relatively high-resolution Pol II ChIP revealed no obvious 5' accumulation. Theoretically, elongation rates put an upper limit on initiation to prevent Pol II traffic jams, even when assuming that every Pol II travels at a constant rate [204]. Surprisingly, this upper limit is almost reached at some transcriptionally active genes, such as *hsp70* [204, 211]. Generally, transcription is not a smooth process: Pol II tends to pause and backtrack throughout transcription [36]. Such fluctuations in the elongation rate could create a more stringent upper limit on how frequently new Pol II can initiate. Elongation could also be rate limiting

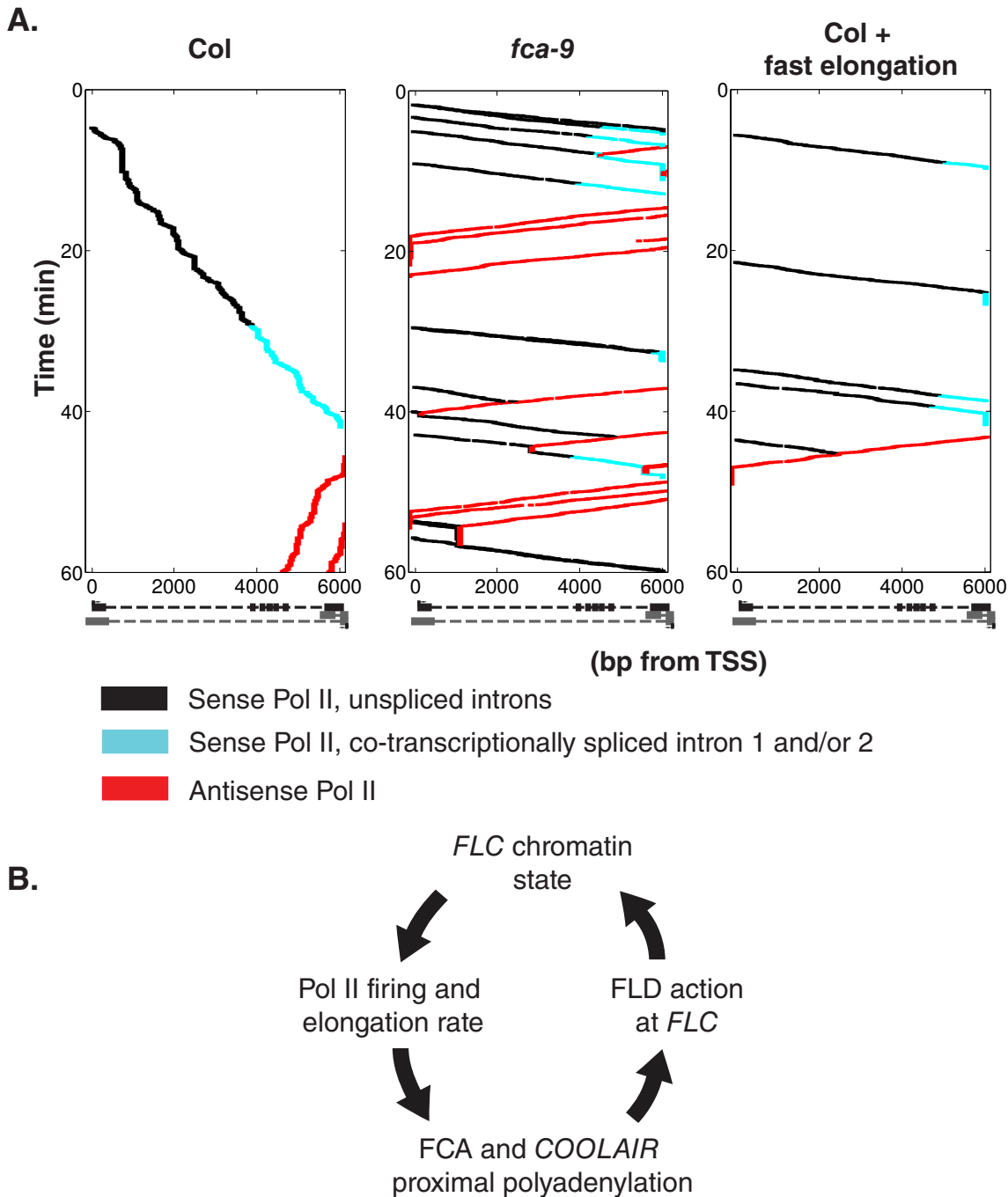


Figure 2.14: Feedback mechanism linking chromatin state to *FLC* repression. (A) Spatiotemporal simulations of Pol II behavior in Col and *fca-9*. Also shown is an outcome of a functional FCA pathway with a constant fast elongation rate (Col + fast elongation). (B) Schematic illustration of how chromatin states at *FLC* coordinate transcriptional initiation and elongation.

during bursts of transcriptional initiation [222]. Furthermore, antisense transcription could induce a limit on initiation rates in order to prevent the occurrence of transcriptional interference between both strands. Coordination of initiation and elongation at *FLC* may have evolved to avoid such transcriptional stress, which has been associated with genome instability [223].

The mechanism of how Pol II initiation and elongation are coordinated remains to be further elucidated. The elongation rate is likely influenced by the chromatin state [206, 224], both directly through nucleosome occupancy, with different histone modifications likely to influence nucleosome wrapping/unwrapping dynamics, and indirectly by affecting the activity of elongation factors. The fast initiation/elongation state at *FLC* is characterized by a series of chromatin modifications, including high levels of H3K4me2/3, H3K36me3, H3Ac with low levels of H3K27me3. SDG8 (EFS), a homolog of ASHH1, is required to establish H3K36me3 genome wide in *Arabidopsis*, including at *FLC* [113]. The low expression *efs* mutant is epistatic to the high expression *fca-9* and *fld-4* mutants [112, 225] suggesting that EFS and H3K36me3 are essential to maintain the high expression state of *FLC* observed in autonomous pathway mutants. This is unlike the situation in yeast where loss of Set2 methyltransferases increases intragenic initiation rather than influences overall expression of the gene [30, 31]. Thus, the changed histone modifications appear to actively influence *FLC* regulation and are not just a reflection of transcription. We still need to fully understand the complexity of the chromatin machinery regulating *FLC* and how the many complexes functionally interconnect [107]. For example, the PAF1 complex is required for high expression of *FLC* [226, 227], while H3Ac also clearly contributes to *FLC* regulation, since inhibition of histone deacetylases results in higher *FLC* expression [127].

Our analysis has investigated how antisense transcription interconnects with chromatin regulation. How *COOLAIR* is directly involved in mediating a repressive chromatin environment remains to be fully established. In subsection 2.2.10 we showed that a model where FLD recruitment (or activity) depends dynamically on proximal *COOLAIR*

polyadenylation in the presence of FCA is consistent with all current data. Further investigations are nevertheless required to dissect how proximal polyadenylation of *COOLAIR* connects to FLD activity. Our stochastic coordination model would predict that slow elongation increases the probability for proximal Pol II termination. In this way slow elongation enhances FLD recruitment indirectly and eventual *FLC* repression (Figure 2.14B). Further experimental evidence will be needed to unequivocally establish this mechanism.

The preponderance of antisense transcription now observed in many genomes opens up the possibility that mechanisms such as the one elaborated here for *FLC* will play widespread roles in gene regulation. An important hypothesis to test using these new tools is whether chromatin-based transcription mechanisms are the major mode of quantitative gene regulation.

## 2.4 Materials and Methods

### 2.4.1 *Determining the Pol II ChIP fragment size distribution*

Using ImageJ (plot profile) we quantified the DNA fragment intensity and the location of fragment length markers along an agarose gel (Figure 2.15A). Combined together, these data reveal the DNA fragment intensity as a function of fragment length. Since the obtained intensity is proportional to the fragment size, we divided the intensity by the fragment length, resulting in an estimate for the relative abundance of fragments, for fragments ranging from 100bp up to 1kb. We then normalized this function to obtain the fragment length distribution (Figure 2.15B).

### 2.4.2 *Convolution of the Pol II density for comparison with experiments*

We need to quantitatively compare the Pol II density arising from the analytical and stochastic models with the experimental Pol II ChIP levels. To facilitate this comparison, we developed a method that in essence generates a predicted Pol II ChIP profile arising from a combination of the underlying model Pol II density and the resolution-limiting effects inherent to the experimental ChIP fragment size dis-

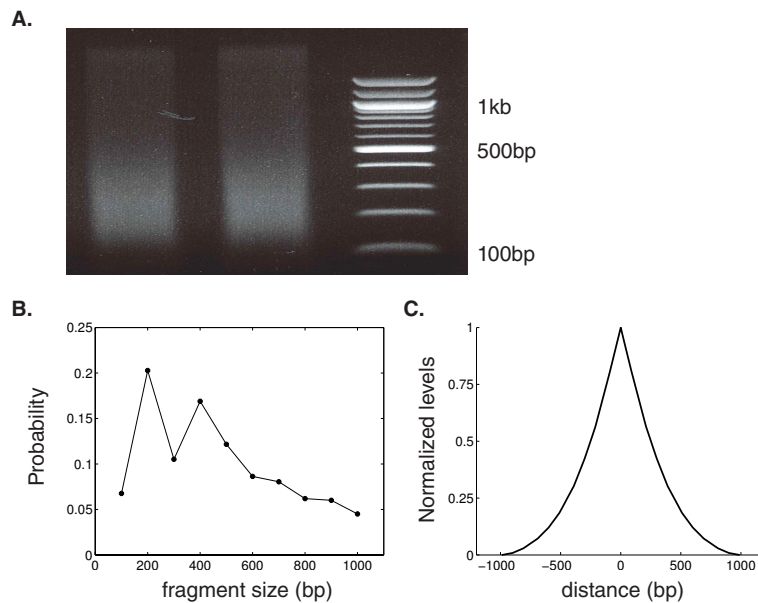


Figure 2.15: Pol II levels obtained in ChIP are affected by the fragment size, related to Figure 2.4. (A) Fragment size intensity in Pol II ChIP assay. The DNA was recovered from sonicated chromatin and was resolved on a gel with DNA size markers running in parallel. (B) Fragment size distribution of lengths ranging between 100bp and 1kb, as extracted from (A). See subsection 2.4.1 for details. (C) Normalized Pol II occupancy resulting from the experimental fragment distribution shown in (B), in presence of a single Pol II located at the origin. Experiments performed by Zhe Wu.

tribution. This procedure is mathematically defined as a convolution and is often used in image processing [228]. Let  $x$  be the position along a gene (in sites of size  $L_{\text{PII}} = 30bp$ ). The convoluted profile  $C(x)$  is proportional to the predicted density of Pol II at position  $x$ , given the uncertainty in fragment length. Intuitively,  $C(x)$  is a sum of all the ChIP fragments that overlap with  $x$  weighted by the number of Pol II bound somewhere along their length. We next define  $I(y)$  as the probability that a ChIP fragment, pulled down with a certain Pol II antibody, is of length  $y$  (in sites). We determined the distribution  $I$  as described in subsection 2.4.1. Let  $y_{\min}$  ( $y_{\max}$ ) be the minimum (maximum) fragment size length. Furthermore, let  $\text{PII}(x)$  be the Pol II density at position  $x$  arising from our theory. This can be used to determine the convoluted Pol II density profile:

$$C(x) = \sum_{y=y_{\min}}^{y_{\max}} \mathbb{P} \left( \begin{array}{c} \text{fragment of} \\ \text{length } y \end{array} \right) \times \sum_{z=(x-y+1)}^x \mathbb{P} \left( \begin{array}{c} \text{fragment 5'} \\ \text{maps to } z \end{array} \right) \text{Number} \left( \begin{array}{c} \text{PII on} \\ \text{fragment} \end{array} \middle| \begin{array}{c} \text{fragment} \\ \text{of length } y \\ \text{5' maps to } z \end{array} \right) \frac{1}{L_{\text{PII}}}$$

$$C(x) = \sum_{y=y_{\min}}^{y_{\max}} I(y) \sum_{z=(x-y+1)}^x \frac{1}{G-y} \sum_{\omega=z}^{z+y-1} \text{PII}(\omega)$$

Here,  $G$  indicates the length of the region from which fragments could arise, in our case on the order of the whole genome. Thus, since  $y_{\max} \ll G$ , the term  $\frac{1}{G-y}$  is effectively a small constant:

$$C(x) \propto \sum_{y=y_{\min}}^{y_{\max}} I(y) \sum_{z=(x-y+1)}^x \sum_{\omega=z}^{z+y-1} \text{PII}(\omega)$$

Using a custom written script in MATLAB, we implemented this formula for both Col and the *fca-9* - mutant Pol II occupancy. Figure 2.4 illustrates the effects of the convolution. The shape of the convoluted profiles are (qualitatively) in line with the experimental Pol II ChIP profiles along the gene (Figure 2.3B,C). The predicted fold upregulation of Pol II is the ratio  $\frac{C^{fca-9}(x)}{C^{\text{Col}}(x)}$  (same for *fld-4*), and can be compared quantitatively with the experimental fold change (Figure 2.4D).

### 2.4.3 Parameter estimation of the co-transcriptional degradation model

Here we detail our procedures for fitting the coTD model to the Pol II ChIP (Figure 2.4D), prior RNA (Figure 2.5) and intronic RNA (Figure 2.6). We include antisense transcription in our analytical model fits to the Pol II data. To determine whether the coTD model could fit the data at least as well as the coordination model, we used the same criteria for fitting as described in subsection 2.2.4. We assume that the targeting rate  $k_{tar}$  is differentially regulated between Col and *fca-9* (or *fld-4*) such that there would be no co-transcriptional decay in *fca-9* (and *fld-4*) ( $k_{tar}^{fca-9} = 0$ ), so we omit the genotype specification for  $k_{tar}$  and  $k_d$  from here onwards. The (simplified) assumption that there is no co-transcriptional decay in the autonomous pathway mutants does not affect our conclusions that arising from the model fitting procedure described below. We allowed  $k_{tar}$  to vary between sense and antisense. We fixed  $k_s = 0.01s^{-1}$  (Col and *fca-9*, *fld*:  $k_s = 0.02s^{-1}$ ), as restricted by our experiments,  $k_{pA,s} = k_{pA,prox} = \frac{1}{50}s^{-1}$  and allowed  $k_{pA,dist}$  to vary, in order to maintain the same criteria with respect to termination as for the coordination model (subsection 2.2.4). Our choices of the termination rate do not affect our conclusions on the model fits below. The nascent RNA degradation rate  $k_d$  and the intron lariat RNA degradation rate  $k_l$  are only relevant in the context of the sense strand. We initially assume no initiation rate changes  $F^{Col} = F^{fca-9}$  so the *a priori* undetermined parameters in this model are:  $\frac{F_{as}^{Col}}{F_s^{Col}}$ ,  $k_{tar,s}$ ,  $k_{tar,as}$ ,  $k_d$ ,  $v^{Col}$ ,  $v^{fca-9}$ ,  $\frac{k_{pA,dist}}{k_{pA,s}}$ . Later we consider a hybrid model that allows initiation rate changes to occur as well and the procedure described below applies to both a coTD and hybrid model.

Recall from subsection 2.2.8 that the model *FLC* mRNA fold upregulation is  $\mathbb{F}(R_s) = \frac{F_s^{fca-9}}{F_s^{Col}} \exp\left(\frac{k_{tar,s}}{v^{Col}} IIA\right)$ . To account for the antisense regulation we assume that degradation can occur up to the distal splice acceptor site with  $IA_{dist}$  the distance from the asTSS.  $3'_{prox}$  denotes the distance between the proximal poly(A) site and the asTSS. Note that by combining proximal Pol II termination (as in the coordination model, detailed in subsection 2.2.4) and coTD in Col we

obtain the following expressions:

$$\mathbb{F}(distal) = \frac{F_{as}^{fca-9}}{F_{as}^{Col}} \frac{1}{\exp\left(-\frac{k_{tar,as}}{\nu^{Col}} IA_{dist}\right)} \frac{\exp\left(-\frac{k_{pA,prox}}{\nu^{fca-9}} L_{PII}\right)}{\exp\left(-\frac{k_{pA,prox}}{\nu^{Col}} L_{PII}\right)}. \quad (2.15)$$

Performing a similar analysis for the proximal antisense, we find:

$$\begin{aligned} \mathbb{F}(proximal) = & \frac{F_{as}^{fca-9}}{F_{as}^{Col}} \frac{1}{\exp\left(-\frac{k_{tar,as}}{\nu^{Col}} 3'_{prox}\right)} \\ & \left(1 - \exp\left(-\frac{k_{pA,prox}}{\nu^{fca-9}} L_{PII}\right)\right) \\ & \times \frac{\frac{k_{pA,prox}}{k_{pA,prox} + k_{tar,as}} \left(1 - \exp\left(-\frac{k_{pA,prox} + k_{tar,as}}{\nu^{Col}} L_{PII}\right)\right)}{\frac{k_{pA,prox}}{k_{pA,prox} + k_{tar,as}} \left(1 - \exp\left(-\frac{k_{pA,prox} + k_{tar,as}}{\nu^{Col}} L_{PII}\right)\right)}. \end{aligned} \quad (2.16)$$

In subsection 2.2.8 we derived the expression for nascent RNA levels in the presence of degradation in Col. In its absence, nascent RNA levels in *fca-9* then reduce to  $F^{fca-9} \left[ \frac{IA-x}{\nu^{fca-9}} + \frac{1}{k_s} \right]$ . The sense model nascent fold changes are thus:

$$\begin{aligned} \mathbb{F}(RNA_{nasc}(x)) = & \frac{F_s^{fca-9}}{F_s^{Col}} \frac{1}{\exp\left(-\frac{k_{tar,s}}{\nu^{Col}} IA\right)} \\ & \times \frac{\left[ \frac{IA-x}{\nu^{fca-9}} + \frac{1}{k_s} \right]}{\left[ \frac{1}{k_{tar,s}} \exp\left(-\frac{k_{tar,s}}{\nu^{Col}} \left(x\left(1 - \frac{\nu^{Col}}{k_d}\right) - IA\right)\right) - \frac{1}{k_{tar,s}} - \frac{x}{k_d} + \frac{1}{k_s} \right]}. \end{aligned} \quad (2.17)$$

Now we proceed with the spatial Pol II distributions of this model including both sense and antisense transcription. *FLC* locus discretization is as described in subsection 2.2.4. Pol II levels in *fca-9* are as described in subsection 2.2.4. It is understood that until this point  $x$  indicated the position relative to *sTSS*. To be explicit, in the remainder of this section  $x \in [0, 208]$  indicates the site number rather than distance with respect to *sTSS*. In the presence of co-transcriptional decay (in Col), we describe for simplicity sense,  $PII_s(x)$ , and antisense,  $PII_{as}(x)$ , Pol II levels at site position  $x$  separately. Total Pol II is still the sum of sense and antisense contributions. Recall the site numbers for sense and antisense TSS: *sTSS* = 4, *asTSS* = 208; termination sites *sTES* = 204 and *asTES<sub>prox</sub>* = 186 and *asTES<sub>dist</sub>* = 0, intron acceptor sites for sense intron1 *IIA* = 131 and distal antisense *IA<sub>dist</sub>* = 12.



We distinguish whether all targeted Pol II have dropped off when reaching the 3' poly(A) site or not for  $sTSS \leq x < sTES$ :

$$PII_s(x) = \begin{cases} \frac{F_s}{v} \exp\left(-\frac{k_{tar,s}}{v}(x - sTSS)L_{PII}\left(1 - \frac{v}{k_d}\right)\right) & \text{if } sTSS \leq x < \min\left(\frac{IA - sTSS}{1 - \frac{v}{k_d}} + sTSS, sTES\right) \\ \frac{F_s}{v} \exp\left(-\frac{k_{tar,s}}{v}(IA - sTSS)L_{PII}\right) & \text{if } \min\left(\frac{IA - sTSS}{1 - \frac{v}{k_d}} + sTSS, sTES\right) \leq x < sTES \end{cases}$$

and at the sense termination site  $x = sTES$ :

$$PII_s(sTES) = \begin{cases} \frac{F_s}{L_{PII}k_{pA,s}} \exp\left(-\frac{k_{tar,s}}{v}(sTES - sTSS)L_{PII}\left(1 - \frac{v}{k_d}\right)\right) & \text{if } sTES < \frac{IA}{1 - \frac{v}{k_d}} + sTSS \\ \frac{F_s}{L_{PII}k_{pA,s}} \exp\left(-\frac{k_{tar,s}}{v}(IA - sTSS)L_{PII}\right) & \text{else.} \end{cases}$$

For the antisense Pol II levels we find upstream of the proximal termination site ( $asTES_{prox} < x \leq asTSS$ ):

$$PII_{as}(x) = \frac{F_{as}}{v} \exp\left(-\frac{k_{tar,as}}{v}(asTSS - x)L_{PII}\left(1 - \frac{v}{k_d}\right)\right).$$

At and downstream of the proximal termination site but upstream of the distal splice acceptor site  $asTES_{dist} < x < asTES_{prox}$ :

$$PII_{as}(x) = \begin{cases} \frac{F_{as}}{v} \exp\left(-\frac{k_{tar,as}}{v}(asTSS - x)L_{PII}\left(1 - \frac{v}{k_d}\right)\right) \exp\left(-\frac{k_{pA,prox}}{v}L_{PII}\right) & \text{if } \max\left(asTSS - \frac{IA_{dist} - asTSS}{1 - \frac{v}{k_d}}, asTES_{dist}\right) < x \leq asTES_{prox} \\ \frac{F_{as}}{v} \exp\left(-\frac{k_{tar,as}}{v}(asTSS - IA_{dist})L_{PII}\right) \exp\left(-\frac{k_{pA,prox}}{v}L_{PII}\right) & \text{if } asTES_{dist} < x \leq asTSS - \frac{IA_{dist} - asTSS}{1 - \frac{v}{k_d}}. \end{cases}$$

and at the distal antisense termination site  $x = asTES_{dist}$ :

$$PII_{as}(asTES_{dist}) = \begin{cases} \frac{F_{as}}{L_{PII}k_{pA,s}} \exp\left(-\frac{k_{tar,as}}{v}(asTSS - asTES_{dist})L_{PII}\left(1 - \frac{v}{k_d}\right)\right) \exp\left(-\frac{k_{pA,prox}}{v}L_{PII}\right) & \text{if } asTSS - \frac{IA_{dist} - asTSS}{1 - \frac{v}{k_d}} < asTES_{dist} \\ \frac{F_{as}}{L_{PII}k_{pA,s}} \exp\left(-\frac{k_{tar,as}}{v}(asTSS - IA_{dist})L_{PII}\right) \exp\left(-\frac{k_{pA,prox}}{v}L_{PII}\right) & \text{else.} \end{cases}$$

The above expressions allow a direct comparison of the model with the experimental data with the criteria as described in subsection 2.2.4. To fit the parameters, we varied  $v^{Col}$  in a range between  $0.3 - 3kb/min$ , consistent with the literature [205]. Given a  $v^{Col}$ ,  $k_{tar,s}$  was then determined by the spliced *FLC* fold change. We then fitted the log of

Equation 2.17 to the log-transformed chromatin-bound *fca-9* fold changes using nonlinear regression. Weighting factors were included as described in subsection 2.2.4. This procedure resulted in good fits ( $R^2 = 0.85 - 0.9$ ) and determined  $\nu_{fold} = \frac{\nu^{fca-9}}{\nu^{Col}}$  and  $k_d$ . It was then possible to also fit the total RNA fold changes by fine tuning  $k_l$ . We then attempted to fit the Pol II ChIP data (Figure 2.3 and Figure 2.4D). At this point only  $\frac{F_{as}^{Col}}{F_s^{Col}}$  and  $\frac{k_{pA, distal}}{k_{pA, s}}$  remained undetermined, so we performed a parameter sweep within the respective ranges described in subsection 2.2.4 to check whether the model could meet the criteria to fit the data. It turned out this was not the case. In subsection 2.2.8 we provided the intuitive reason behind this result.

In our analysis above we have assumed that  $k_d > \nu$  so that a fraction of the Pol II would terminate under the influence of Xrn. In a scenario where  $\nu \geq k_d$  the  $RNA_t$  would be larger relative to the other RNA fractions. This would be inconsistent with chromatin-bound RNA fold changes, because the fitting procedure indicated it was required that  $k_d \gg \nu$ . We conclude that a coTD model is not consistent with the available experimental data despite having more *a priori* undetermined parameters than the coordination model.

Parameter	Description	Value	Notes
$L_{PII}$	Length of a Pol II site along the <i>FLC</i> gene	30bp	This length corresponds to the size of a Pol II footprint [219].
$L$	Length of <i>FLC</i> gene in sites	209	For simulation purposes <i>FLC</i> is divided into $L$ sites of length $L_{PII}$ .  At most one Pol II can reside at each site.
$P_{total}$	Total Pol II in a simulation	1000	We assume a fixed number of Pol II inside a nucleus throughout the simulations, i.e. $P_{total} = P_{free} + \text{Pol II transcribing } FLC$ .
$H$	Number of histone H3 at <i>FLC</i>	70	As in [138], the A mark levels range from 0, 1, ..., $H - 1$ , $H$ .
$T_{elon}$	Elongation threshold	11.8	Fitted to Pol II levels. Indicates the A level above which Pol II can elongate.
$k_{elon}$	Elongation parameter	$0.018s^{-1}$	Fitted to Pol II levels, determines the elongation rate.
$k_s$	Splicing (cleavage) rate	$0.01s^{-1}$ (Col, <i>fca-9</i> ) $0.02s^{-1}$ ( <i>fld-4</i> )	Constrained by experiment [44] and our measurements of splicing efficiency Figure 2.6D.
$k_l$	Intronic RNA degradation rate	$0.05s^{-1}$	Constrained by experiment [44], 5' to 3' degradation assumed in line with (Hesselberth, 2013), corresponds to $1.5bps^{-1}$ .
$k_{pA}$	Pol II termination rate at poly(A) site	$0.02s^{-1}$	Constrained by experiment [41]. This value primarily determines our absolute mean initiation and elongation rates.
$k_{col}$	Pol II termination rate due to sense/ antisense Pol II collision	$0.02s^{-1}$	Taken to be the same value as $k_{pA}$ in absence of further information.
$T_{f,s}$	Sense initiation threshold	14.8	Fitted to spliced <i>FLC</i> fold changes and Pol II levels.
$k_{f,s}$	Sense initiation parameter	$8 \cdot 10^{-8}s^{-1}$	Fitted to spliced <i>FLC</i> fold changes and Pol II levels.
$T_{f,as}$	Antisense initiation threshold	10.5	Fitted to proximal/distal as fold changes and Pol II levels.
$k_{f,as}$	Antisense initiation parameter	$5 \cdot 10^{-8}s^{-1}$	Fitted to spliced <i>FLC</i> fold changes and Pol II levels.
$k_{FLD}$	FLD inactivation rate	$10^{-5}s^{-1}$	This parameter, together with the A mark levels determined by $k_A$ , $k_{r,0}$ and $k_{r,s}$ has to generate dynamics slow enough so that low A mark levels persist over timescales long enough to ensure continuous repression.
$k_A$	Constitutive A placement rate	$4.4 \cdot 10^{-3}s^{-1}$	Fitted to experimental H3K36me3 fold changes.
$k_{r,0}$	Constitutive A removal rate	$1.7 \cdot 10^{-3}s^{-1}$	Constrained by experiment [220].
$k_{r,s}$	FLD-stimulated A removal rate	$7.5 \cdot 10^{-3}s^{-1}$	Fitted to experimental H3K36me3 fold changes and Pol II levels.
$k_x$	mRNA export rate	$1.7 \cdot 10^{-3}s^{-1}$	Constrained by experiment [41].
	Col + fast elongation simulations		All parameters remain unchanged unless indicated below.  $T_{elon}$ is not present in this model.
$k_{elon}$	elongation rate	$0.9s^{-1}$	Fitted such that the mean elongation rates are equal to 1.5kb/min. The elongation rate is now independent of the A mark.

Table 2.1: Parameter values used in the stochastic coordination model



## VARIATION IN *FLC* TRANSCRIPTION, PROCESSING AND DEGRADATION

---

### 3.1 Introduction

Up to this point we have investigated *FLC* regulation on a level of whole plant cell population averages. This line of research has been fruitful. However, it remains unclear to what extent cellular variation in *FLC* expression is present in *Arabidopsis*. Spatial gene expression patterning has long been recognized to be of importance throughout the development of plants and other multicellular organisms [90, 229–233]. In recent years, it has become clear that stochasticity is another major contributing factor to cellular variation in gene expression in organisms ranging from microbes to humans [144, 234, 235]. It still remains unclear whether this is also the case in plants and more specifically for *FLC* regulation.

Stochasticity in gene expression can have beneficial or harmful consequences to organisms. It is observed in a variety of situations, including stress response, metabolism, development, the cell cycle, circadian rhythms, and aging [234]. Cell-to-cell variation can be classified as intrinsic or extrinsic [236, 237]. The inherent stochasticity of biochemical processes such as transcription and translation generates intrinsic noise [237]. In addition, fluctuations in the amounts or states of other cellular components lead indirectly to variation in the expression of a particular gene and thus represent extrinsic noise. Extrinsic variation can also occur when expression is dependent on a factor that differs systematically between cells, for instance cell volume [235]. It has been shown theoretically that when translation leads to production of already more than  $\sim 2$  proteins per mRNA, intrinsic noise is dominated by transcription [237]. Interestingly, transcription can occur in so called 'bursts', where in a short period of time multiple transcripts are made, followed by a relatively long period where transcription activity is low or absent [238–240]. Such bursting can certainly contribute to gene expression variation and

is usually associated with changes in transcriptional initiation [237, 239], although Pol II pausing and backtracking during elongation could also contribute to burstiness [222]. The gene regulatory network structure controlling a particular gene, can also contribute to the stochasticity of that particular gene's expression [144, 241].

In this chapter we investigate cellular variability in *FLC* expression and its implications for regulation by the autonomous pathway and the transcriptional activator *FRIGIDA* (*FRI*) in non-vernalized conditions. We build upon our findings in the previous chapter, that described regulation in terms of cell population averages in whole plant seedlings. By combining theory from stochastic processes and single molecule experiments in plants with high *FLC* expression due to the presence of *FRI*, we find that *FLC* processing and degradation are quantitatively well described by Poisson processes. *FLC* transcription scales strongly with cell volume, which explains the large cellular variability. We further confirm that *FLC* transcription is not 'bursty' by quantitative fluorescence image analysis. Furthermore, we estimate *FLC* and *COOLAIR* transcription initiation and elongation rates and then use these rate estimates to conclude that coordination of transcription initiation and elongation not only occurs between Col and autonomous pathway mutants, but also between *Arabidopsis* plants with and without an active *FRI* allele. We predict that fast elongation rates could be functionally important to achieve a high *FLC* expression state. Altogether, we propose that *FLC* expression is tightly maintained at a certain cellular concentration by *FRI* and the autonomous pathway to ensure a quantitative tuning of *Arabidopsis* flowering time.

## 3.2 Results

### 3.2.1 *Cellular FLC distribution is monomodal and super-Poissonian*

To investigate whether there is cellular *FLC* variation in ColFRI (Columbia accession with a homozygous active *FRI* allele), we utilized a technique termed single molecule RNA fluorescence in situ hybridization (smFISH). This method allows simultaneous detection, localization and quantification of individual RNA molecules at the sub-

cellular level in fixed samples and has been established for yeast and mammalian cells [144, 235, 242]. smFISH is based on multiple single-labeled oligonucleotide probes binding to a target RNA ([242], Figure 3.1A). Due to the high number of probes binding to the same RNA target, this generates a diffraction-limited signal that can be detected by fluorescence microscopy (Figure 3.1A,B). In *Arabidopsis* the protocol to image root cells has recently been developed by Stefanie Rosa and Susan Duncan [243]. Sample preparation first involves squashing a root. Cells can then be fixated, treated with the RNA probes and further imaged (Figure 3.1A). In our experiments, we used two distinct probe sets simultaneously: one covering sense *FLC* exons (exon RNA) and the second covering sense intron1 (unspliced RNA) (Figure 3.1A). We also used the DNA stain DAPI to label the nucleus. Quantification of cellular RNA in roots was achieved by combining an automated spot detection image analysis with a computational segmentation of the fluorescence signal to determine the cell outlines (method developed by Matthew Hartley and Tjelvar Olsson, [243]). We calibrated the computational spot detection analysis method based on the size and shape of the diffraction-limited spots on a test set to ensure the estimated exonic RNA quantities from the algorithm matched the manual estimations. For the unspliced RNA we manually inspected all samples. Throughout our analysis we made the widely accepted assumption in the field that one diffraction-limited cytoplasmic exon RNA spot represents one RNA molecule [242]. For further information on details of the smFISH method in non-plant systems we refer to [242]. For the smFISH protocol specific to plants and automated image analysis algorithm we refer to [243].

We observed that exon RNA was mostly cytoplasmic and abundant, indicating that *FLC* mRNA can be readily detected with the exonic probes (Figure 3.1B). The unspliced RNA was only detected in the nucleus (Figure 3.1B) with counts ranging from 0 up to 4 per cell (Figure 3.2A). Since the observed root cells (in the meristematic region) are not expected to be endoreduplicated, these spots likely indicate the *FLC* loci with splicing occurring at the locus as the gene copy number increases from 2 to 4 throughout the cell cycle. The

cellular distribution of exon *FLC* ( $N_{cells} = 438$ ) is monomodal with a mean ( $\pm$ s.e.m.) of  $59 \pm 2$  molecules (Figure 3.1C). This distribution is a pool of 8 roots arising from 3 different experiments. Variation between replicates is present with means ranging from 43 to 71, however in each replicate the wide variation in cellular RNA counts was observed. The measurement error in spot detection algorithm does not exceed this biological variation. In the absence of a positive control we cannot completely rule out that we do not detect all RNAs present in the cell, however that would only lead to systematic underestimation of the RNA quantities. Unless stated otherwise, this caveat does not affect our conclusions as long as the fraction of unobserved RNAs remains constant between samples.

By comparing the exon RNA levels to a Poisson distribution (Figure 3.1C), we found that the RNA distribution is super-Poissonian (KS test,  $p = 3 \cdot 10^{-51}$ ): its variance is  $\sim 16$ x larger than the mean of the distribution. A Poisson distribution would be expected if the RNA levels are determined by a birth-death process. This continuous-time Markov process counts the number of RNAs and is a generalization of a deterministic situation of a constant RNA production and first order degradation rate. A birth-death process assumes that RNA production occurs as a Poisson process: production events occur (1) independently of each other and (2) with constant probability per unit time  $p$  [244–246]. The same conditions (1) and (2) are imposed on degradation with a constant degradation probability per unit time  $d$ . Even though RNA degradation does not strictly satisfy the mathematical definition of a Poisson process [244], we still refer to it here as a Poisson process because it does satisfy the above characteristic features (see subsection 3.4.1 for more details). To illustrate how to interpret these conditions: if RNA would be produced in transcriptional 'bursts', assumption (1) would break down. In a scenario where the production rate is time-dependent (see subsection 3.2.3), assumption (2) would no longer be true. In those cases the expression variation can be broader and even bimodal [222, 238–240]. In subsection 3.4.1 we show that this conclusion is not dependent on the (simplest) assumption that RNA production and degradation are modeled as single step processes (one stage birth-death



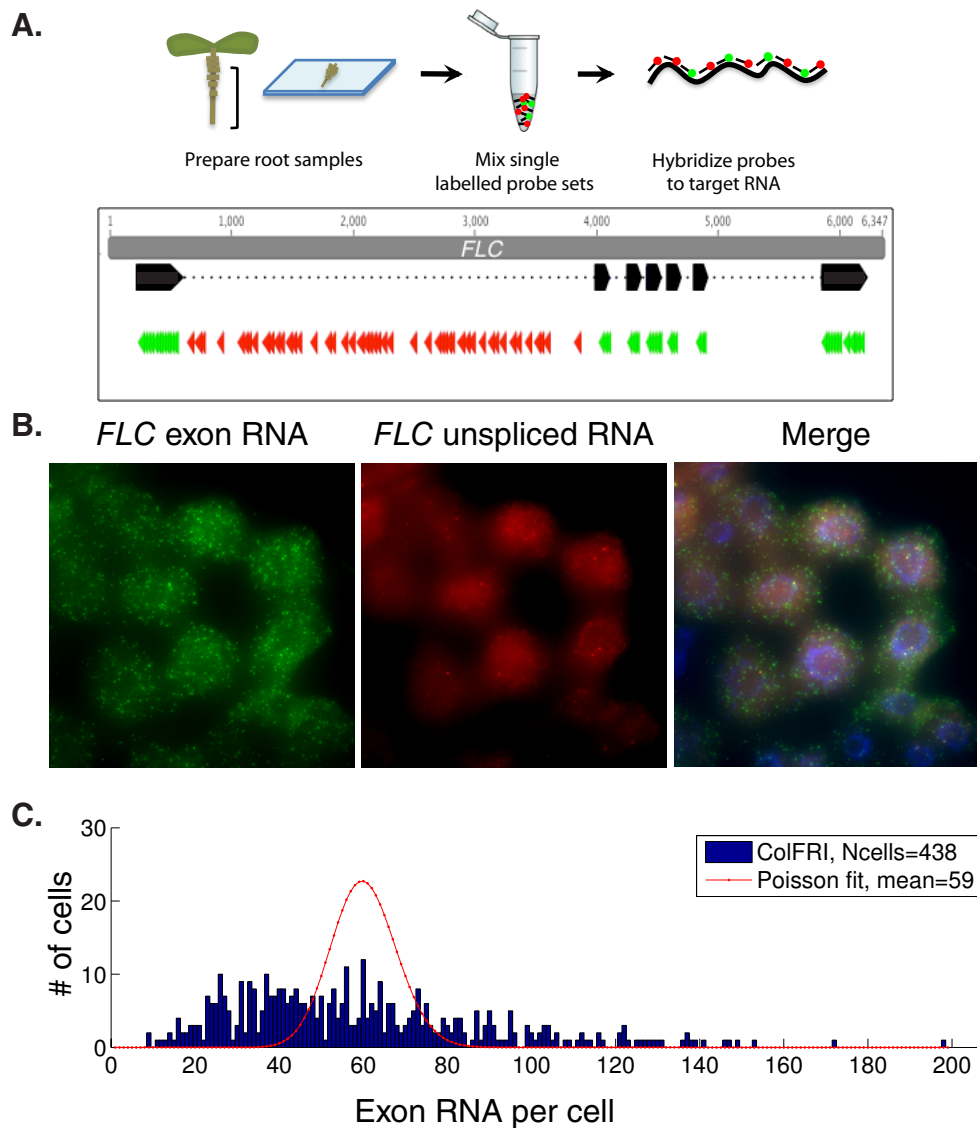


Figure 3.1: (A) Schematic of smFISH protocol in plants as described in [243] and schematic of *FLC* gene with different probe sets for labelling sense exon RNA (green) and sense unspliced RNA (red). (B) Fluorescence localization of exon RNA (green), unspliced RNA (red) and a merge containing exon RNA, unspliced RNA and DAPI stain (blue) in representative *Arabidopsis* root cells. (C) Histograms of cellular exon RNA counts (blue) and a Poisson distribution fit (red). Images (A) and (B) are a courtesy of Stefanie Rosa and Susan Duncan who performed the experiments.

process). There, by considering a nascent RNA species that can be processed into a mature RNA species (two stage birth-death process), we show that both the nascent and mature RNA distributions would still be Poissonian. This result can be extended to multiple intermediate RNAs. So we conclude that either assumption (1) or (2), underlying the Poisson distribution, is not true for *FLC* exon RNA production or degradation. Below we further investigate what aspects of regulation cause an enhanced variation in *FLC* expression.

### 3.2.2 *FLC degradation satisfies Poisson characteristics*

To determine if *FLC* exon RNA degradation is responsible for the observed super-Poissonian variation in *FLC* in expression, we hypothesized what might happen to the distributions after transcription were inhibited. Intuitively, at first the RNA distribution would be as observed in Figure 3.1C, but as time progresses we expect the RNA levels to drop. From the mean RNA levels  $\langle R(t) \rangle$  over time, the (mean) degradation rate  $d$  (units:  $s^{-1}$ ) is often calculated by fitting these to an exponential function  $\langle R(t) \rangle = \langle R(0) \rangle \exp(-dt)$  [244, 247, 248]. This function arises under the assumption that RNA degradation is well described by a first order reaction. The stochastic process describing a first order degradation reaction would be a pure death process  $\{R(t)\}_{t \geq 0}$  characterized by the constant probability per unit time  $d$  and the Poisson criteria as described above [244]. To test whether RNA degradation satisfies these Poisson characteristics, we derived the expected probability distribution as a function of time  $\mathbb{P}(r, t) := \mathbb{P}(R(t) = r)$  given the observed initial distribution  $\mathbb{P}(r, 0)$  before transcription inhibition (subsection 3.4.1). This resulted in:

$$\mathbb{P}(r, t) = \sum_{b=r}^M \binom{b}{r} (\exp(-dt))^r (1 - \exp(-dt))^{b-r} \mathbb{P}(b, 0), \quad (3.1)$$

where  $M := \max\{r : \mathbb{P}(r, 0) > 0\}$  indicates the maximum amount of cellular RNA observed in the initial distribution. This result can also be obtained in a more intuitive manner. For each RNA that was present initially the probability that it has not been degraded yet after a given time  $t$  follows an exponential distribution and is thus equal to  $\exp(-dt)$ . The probability to still have  $r$  RNAs left after time  $t$  out of  $b \geq r$  RNA initially is then binomially distributed with parameter  $\exp(-dt)$ . The resulting binomial probability multiplied by the prob-

ability to have  $b$  RNA initially and summed over all possible  $b$  then also gives the desired result.

We then tested this hypothesis experimentally by treating plant seedlings with Actinomycin D (ActD) and subsequently performing smFISH in a time series. ActD is well established as a transcription elongation inhibitor in plants and other organisms [247, 249]. We observed that unspliced RNA signal was almost completely gone compared to a mock (DMSO) treated sample after 4 and 6 hours of ActD addition, indicating that transcription inhibition was successful (Figure 3.2A).

We used the mean exonic RNA levels after 4 and 6h treatment to estimate  $d$ :  $\frac{\langle R(4h) \rangle}{\langle R(6h) \rangle} = \exp(d(6h - 4h))$ . This resulted in  $d = 3.25 \cdot 10^{-5} s^{-1}$ , equivalent to a half life of  $t_{1/2} = \frac{\ln(2)}{d} = 5.9h$ . We then compared the model distributions (Equation 3.1) with the experimentally observed RNA distributions after 4 and 6 hours (Figure 3.2), by taking the pooled mock treated RNA distributions as initial distributions in the model to minimize the influence of variation between experiments (Figure 3.2B). The variation in the mock treated distribution resembled the untreated distribution well, but the former had slightly higher mean levels (68 as compared to 59 in Figure 3.1C). The cumulative distribution functions of the model and experimental levels were similar, as shown in Figure 3.2B (KS test,  $p = 0.66$  and  $0.21$  for 4h and 6h respectively). This good fit is under the assumption that degradation started after a time lag  $\tau$  of 1h after ActD treatment (through replacing  $t$  by  $t - \tau$  in Equation 3.1 for  $t = 4, 6h$ ). Such a time lag is not unreasonable considering the ActD penetration time into the plant tissue. We also performed the same procedure using the 4h distribution as the initial condition. There we could correctly predict the 6h distribution with the same degradation rate and no time lag (KS test,  $p = 0.48$ ). We conclude that *FLC* exonic RNA degradation is well described by a Poisson process with a mean half life of 5.9h and that the observed super-Poissonian distribution (Figure 3.1) is not caused by fluctuations in the degradation process.

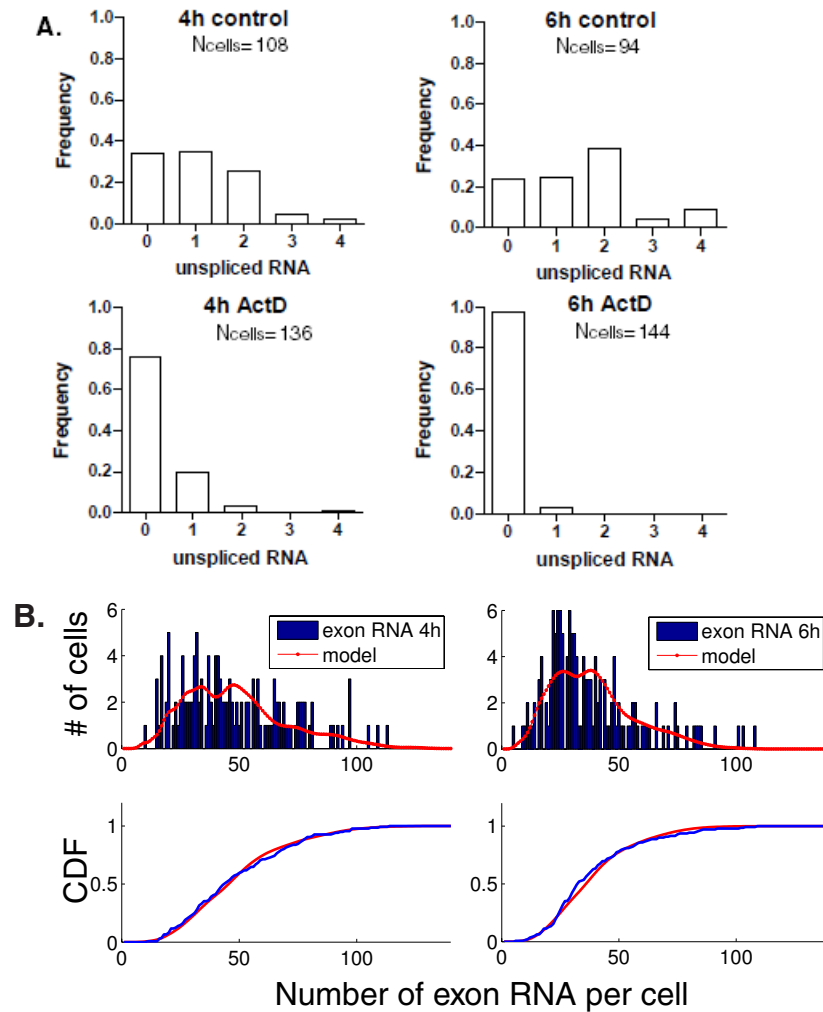


Figure 3.2: (A) Histograms of cellular unspliced RNA counts for 4 and 6h after ActD or mock treatment (control). After 4h almost all unspliced signal has disappeared. (B) Histograms of cellular exon RNA counts (blue) and model fit (red). Numbers of cells in the exon RNA counts are as for 4h and 6h ActD treatment. In the bottom panels the cumulative distribution functions (CDFs) are displayed for both the exon RNA counts and model fits. Experiments performed by Stefanie Rosa, Zhe Wu.

### 3.2.3 *A constant transcription rate per locus is not sufficient to explain FLC variation*

We proceeded with investigating potential sources of variation in *FLC* transcription. In *Arabidopsis* the cell cycle for meristematic root cells (the cells analysed in our smFISH assay) is on average 17h [250, 251]. In these cells *FLC* copy number increases from 2 in G1 to 4 by the end of S phase and during G2. During mitosis, transcription seems not to occur so that the copy number is effectively zero [243]. Together with the (simplest) assumption that each locus produces RNA with a constant rate  $p$ , this copy number variation could give rise to a time-dependence in the overall (mean) RNA production rate. To investigate this hypothesis further we developed an ODE model to describe the temporal dynamics for the (mean) exon RNA levels  $R(t)$  as a function of time in the cell cycle. We used cell cycle stage time period estimates determined by [250, 251]. Let  $N_{loci}(t)$  represent the copy number throughout the cell cycle for  $0 \leq t \leq 17h$ :

$$N_{loci}(t) = \begin{cases} 2 & \text{if } 0 \leq t < 7h, \\ 3 & \text{if } 7h \leq t < 8h, \\ 4 & \text{if } 8h \leq t < 14.5h \\ 0 & \text{if } 14.5h \leq t \leq 17h. \end{cases} \quad (3.2)$$

Then the ODE can be described as:

$$\frac{dR(t)}{dt} = pN_{loci}(t) - dR(t). \quad (3.3)$$

To obtain the solution  $R(t)$  we solved the ODE piecewise analytically. To ensure continuity we assume that upon cell division, the RNA levels are halved (see subsection 3.4.2 for further details).  $R(t)$  is a function of two parameters:  $p$  (*a priori* unknown) and  $d$  (estimated in the previous section). Note that in this model steady state is reached exponentially quickly (see Equation 3.9). However, in our particular solution, RNA levels are never constant and vary  $\sim 3x$  throughout the cell cycle (Figure 3.3A). These features arise because the associated time scale to reach steady state is dictated by the relatively long *FLC* half life of 5.9h (see subsection 3.2.2). By equating the *experimental* mean exonic RNA levels  $\langle R \rangle_{exp}$  with the model cell cycle average  $\langle R(t) \rangle_{T_{cell}} = \frac{1}{T_{cell}} \int_0^{T_{cell}} R(\tau) d\tau$ , we can estimate the production rate per locus to be  $p = \frac{1}{17} \text{min}^{-1}$ .

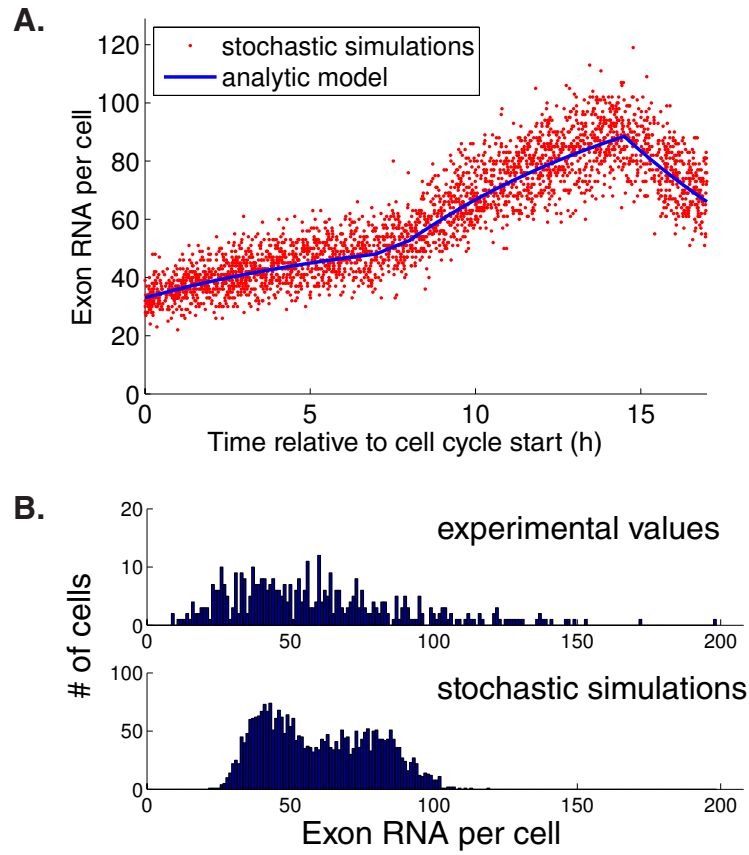


Figure 3.3: (A) Analytical  $R(t)$  solution to ODE model as a described in Equation 3.3, indicating mean exon RNA levels per cell as a function of time (blue) and exon RNA levels, sampled at random time points during the cell cycle (red) arising from stochastic birth-death simulations ( $N_{cells} = 3000$ ). (B) Histograms of cellular exon RNA levels: experimental values (top, as in figure Figure 3.1C) and arising from stochastic model simulations (bottom) described in (A).

Now that we have estimated this parameter  $p$ , we can test whether the observed *FLC* variation could be explained by a birth-death process with a time-dependent (mean) production rate  $pN_{loci}(t)$  and (mean) degradation rate  $d$ , through stochastic simulations. As before we assume stochastic events occur independently of each other. We implemented a Gillespie algorithm [218] in C++ with the reactions and corresponding propensities  $p_t$  described in Table 3.1. Parameter values used are listed in Table 3.2. After every  $T_{cell}$  we halved the RNA levels. This is justified because of the relatively high copy number of *FLC* mRNA. Furthermore in this way we also ensure that variation in *FLC* levels in the model only arise from RNA production and degradation.

To compare the model with the experimental *FLC* distribution we

Reactions	Propensities $p_t$
$\emptyset \rightarrow R$	$p \cdot N_{loci}$
$R \rightarrow \emptyset$	$d \cdot R$

Table 3.1: Reactions and propensities used in the birth-death model with varying *FLC* copy number

Parameter	Description	Value	Notes
$p$	production probability per unit time per locus	$10^{-3} s^{-1}$	Constrained by experiment, see subsection 3.2.3.
$N_{loci}$	<i>FLC</i> copy number	2 if $0 \leq t < 7h$ , 3 if $7h \leq t < 8h$ , 4 if $8h \leq t < 14.5h$ 0 if $14.5h \leq t \leq 17h$ .	Constrained by experiment [250, 251], see subsection 3.2.3
$d$	degradation probability per unit time	$3.25 \cdot 10^{-5} s^{-1}$	Constrained by experiment, see subsection 3.2.2.

Table 3.2: Parameter values used in the birth-death model with varying *FLC* copy number

sampled RNA levels at randomly chosen time points throughout the cell cycle (Figure 3.3A). To achieve this we sampled (using a uniform distribution) a time point  $T$  between 0 and  $T_{cell}$ . We started

without any initial RNA and then ran the simulation for duration  $100T_{cell} + T$  (in simulated time). The amount of *FLC* present at the end of the simulation provided the sampled RNA value corresponding to time point  $T$ . We performed 3000 simulations to give rise to the temporal variation shown in Figure 3.3A. The histogram of all sampled model RNAs (Figure 3.3B) is more variable than a Poisson distribution ( $\lambda = 60$ , KS test  $p = 5 \cdot 10^{-228}$ ), but still less variable than the experimental one (KS test  $p = 3 \cdot 10^{-9}$ ). We conclude that copy number variation throughout the cell cycle could contribute to *FLC* variation, but together with a constant transcription rate per locus, this is not sufficient to explain the observed *FLC* variation.

#### 3.2.4 *A transcription rate scaling with cell volume can explain FLC variation*

Recently it has been shown that cellular RNA levels can correlate with cell volume in mammals [235]. Visual inspection suggested this might also be the case for *FLC* in plant cells (Figure 3.1). We performed a systematic unbiased analysis to quantify this effect. By extending our computational segmentation algorithm to output the cell area in pixels, we compared this with the corresponding exon RNA counts (Figure 3.4A). We found a strong linear correlation between cell area and RNA levels ( $R^2 = 0.91$ ). The intercept  $\alpha = 6 \pm 2$  molecules ( $p = 4 \cdot 10^{-3}$ ) of the linear fit (F-statistic:  $p = 2 \cdot 10^{-66}$ , with slope  $\beta = 3.7 \pm 0.1 \cdot 10^{-3} \text{ pixel}^{-1}$ ) suggested that per cell only a very small fraction is present independently of cell area. *Arabidopsis* roots are radially symmetric and the depth of the cells along the radial axis is relatively constant between cell layers [252]. So our cell area estimate can be considered a proxy for cell volume. Given that *FLC* degradation is well described by a Poisson process with a (volume-independent) constant degradation rate  $d$  (subsection 3.2.2), we conclude that the *FLC* production rate needs to scale with volume in order to explain the scaling of exon RNA levels with volume.

Interestingly, we did not find a correlation ( $R^2 = 0.02$ , F-statistic:  $p = 0.12$ ) between cell area and the number of unspliced RNA per cell (Figure 3.4B). This is not necessarily inconsistent with a transcription rate scaling with volume because the number of unspliced



RNA is also a function of *FLC* copy number. Although cell cycle stage might have a correlation with cell volume, given our observed large variation in area ( $\sim 10\times$ ), we consider it highly likely that small cells can be in any stage of the cell cycle. A proportion of small cells will then be in G2, so that 4 unspliced spots could be observed, despite their small cell volume. Likewise, bigger cells also spend a fraction of the time in G1, so that only 2 spots can be observed. These restrictions might mask a correlation between the unspliced signal and volume.

To disentangle how much of the variation in *FLC* is governed by

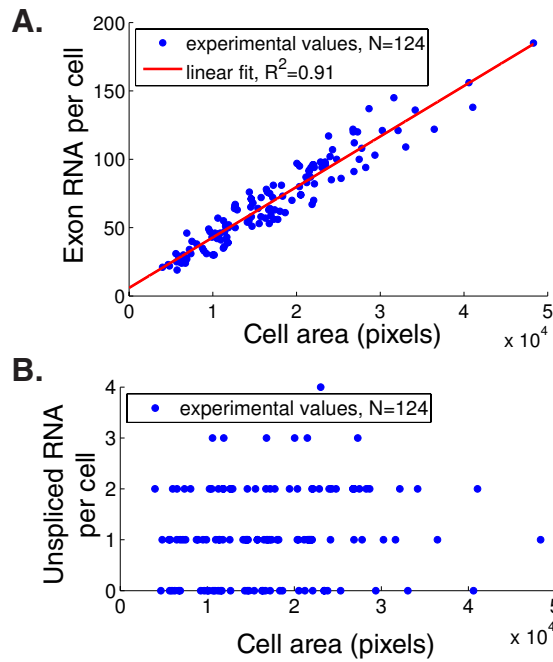


Figure 3.4: (A) Scatter plot of cellular exon RNA (blue) as a function of the cell area (pixels) together with a least squares fit (red). (B) Scatter plot of cellular unspliced RNA as a function of cell cycle, no significant correlation is observed. Experiments performed by Stefanie Rosa, Zhe Wu.

the cell size and how much due to other sources of variability, we can consider the cellular exon RNA levels as a random variable  $R(A)$  such that its expectation given the cell area  $A$  is given by the observed linear relationship [235]:  $\mathbb{E}[R|A] = \alpha + \beta A$ . Here,  $\alpha$  and  $\beta$  are the respective intercept and slope of the fit in Figure 3.4A. A volume cor-

rected noise measure can be defined as  $N_R := \frac{\text{Var}[R] - \text{Var}[R|A]}{\mathbb{E}[R]^2}$ . In [235], it was shown that  $N_R$  can be expressed as:

$$N_R = \frac{\text{Var}[R]}{\mathbb{E}[R]^2} - \frac{\beta \mathbb{E}[A]}{\alpha + \beta \mathbb{E}[A]} \frac{\text{Cov}[R, A]}{\mathbb{E}[R] \mathbb{E}[A]}$$

The right hand side can be readily calculated using the experimental data. If volume is the major source of variation, then we expect  $N_R$  to be small. In fact, if volume is the major determinant of the variation, then we expect the residual variation to be caused merely by (intrinsic) Poisson fluctuations:  $\mathbb{P}[R|A] \sim \text{Pois}(\lambda = \alpha + \beta A)$ . This is the so called Poisson limit [235]. In that case it can be shown that:  $N_R = \frac{1}{\mathbb{E}[R]}$ . We estimated the volume corrected noise measure for our *FLC* levels and indeed finding  $N_R = 0.017$ , very close to the Poisson limit  $\frac{1}{\langle R_e \rangle} = 0.014$ . This is strong evidence in support of the notion that, for given a cell volume, *FLC* transcriptional regulation in ColFRI can be described by a Poisson process. As explained in subsection 3.2.1, a Poisson transcription process contrasts with a scenario where RNA is produced in transcriptional 'bursts', as the latter is characterized by a larger variation in RNA levels.

### 3.2.5 *FLC* RNA accumulates little at the locus

A feature of transcriptional bursting is the accumulation of multiple RNAs at a locus [239, 240]. To further investigate whether sense *FLC* transcription occurs in bursts or uncorrelated events, we performed quantitative image analysis on the exon spot intensities [253] to see if there is RNA accumulation at *FLC* loci. We extended our computational algorithm that determined diffraction-limited spots [243] to also calculate spot intensity values as the sum of the pixel values that lie within the spots. Accumulation of multiple RNAs should result in higher intensity values, assuming there is no saturation of the signal, and a larger spot size [254]. However, manual inspection suggested that there were no large spots visible in our sample (see Figure 3.1B for representative examples). An unbiased systematic acquisition of all spots resulted in a distribution of exon RNA intensities (Figure 3.5A). The intensity variation is most likely affected by different numbers of probes binding to cytoplasmic RNAs. Compared to the median of the distribution ( $median = 6.5 \cdot 10^5$  in arbitrary units (A.U.)), the maximal values were only 2 – 3x higher

( $max = 19 \cdot 10^5$  (A.U.)). So under the assumptions that the median value (Figure 3.5A) resembles the average intensity of a single RNA, and that the variation is completely generated by multiple RNAs accumulating, there would still only be accumulation of 2–3 molecules in rare events. It could be that the intensity values are suffering from a high background, such that the variation due to differences in RNA accumulation are not transferred into a linear scaling of intensity values. A control experiment that can determine whether the intensity scales with number of RNAs and/or probes, would be beneficial to dissect this issue. However given that the size of the spots do not change, we currently do not favour this scenario.

Unspliced RNA spots likely indicate *FLC* loci (subsection 3.2.1). We identified which exon RNA spots (if any) were within a 3 pixel proximity to these unspliced RNA locations. Compared to the median intensity of all exon RNAs, a range of intensities was observed for these locus-associated exon RNAs (Figure 3.5B). A large proportion of the loci did not have a spot associated, indicated in the histogram with intensity 0 (Figure 3.5B). We conclude that at *FLC* loci, likely at most  $\sim 3$  exon RNAs are present, but these cases seem rare. It could also be that these high intensity cases reflect the clustering of *FLC* loci [255], so that the observed RNA accumulation is not due to an increased transcriptional activity per locus. For the overall majority of loci, only 0 or 1 exon RNA seem to be present. Unlike for exon RNA, we lack a benchmark intensity value for the unspliced RNA to indicate a single RNA. However, we observed little intensity variation, which might suggest that unspliced RNA would not significantly accumulate at the locus either (Figure 3.5C). We conclude that our intensity analysis provides some evidence that *FLC* transcription does not occur in bursts within a time scale of exon RNA export from the locus, which could be on the order of  $\sim 10min$  as found in other systems [41]. Further conclusive evidence is needed to unequivocally determine the number of RNAs accumulating at the locus and whether the times between *FLC* transcription events are really uncorrelated.

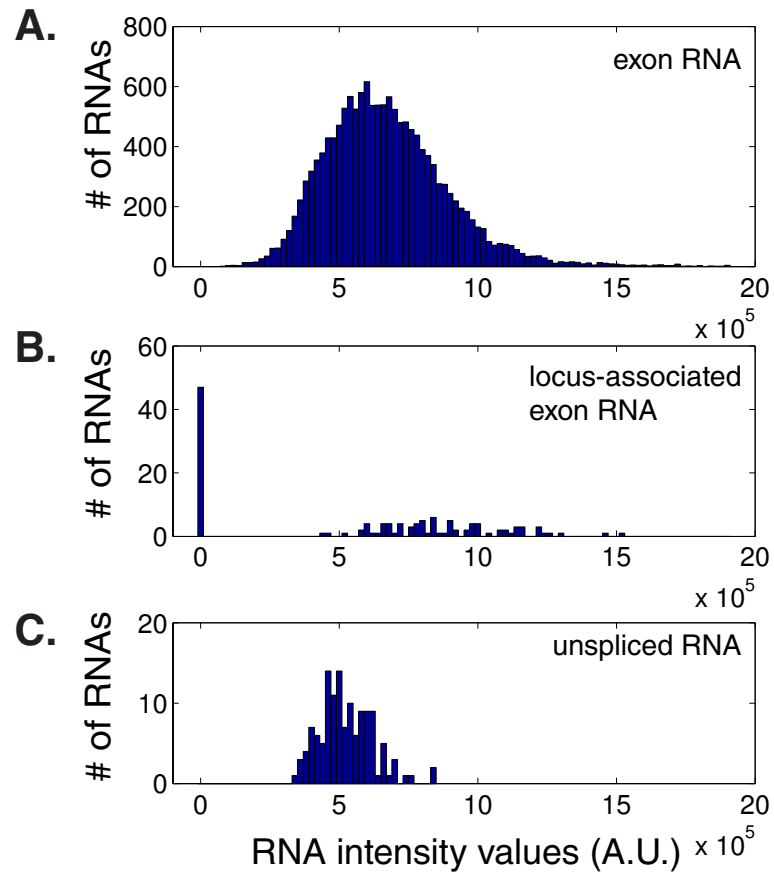


Figure 3.5: Intensity values for cellular exon RNA (A), locus-associated exon RNA (B) and unspliced RNA (C). The median value for exon RNA is  $median = 6.5 \cdot 10^5$  arbitrary units (A.U.) with maximal value  $max = 19 \cdot 10^5$ . Experiments performed by Stefanie Rosa, Zhe Wu.

### 3.2.6 Estimation of *FLC* and *COOLAIR* production and splicing rates

Since we have determined absolute cellular RNA levels, we can utilize this data to infer absolute RNA production rates. This is not possible with methods that rely on amplification of signals such as RT-qPCR or RNA-seq. Absolute rate estimates are useful to dissect regulatory mechanisms that occur only in certain parameter regimes, for instance transcriptional interference or traffic jams in case of high Pol II initiation events. We will investigate these features in the following sections. Recall from chapter 2 that (assuming steady state) population averaged RNA levels are a ratio of RNA production and degradation  $\langle R \rangle = \frac{p}{d}$ . In this expression the *FLC* copy number is implicitly incorporated in the value of  $p$ . We have determined the degradation rate  $d$  experimentally (subsection 3.2.2). Here, we consider a cell cycle averaged copy number arising from Equation 3.2:  $N_{loci} := \langle N_{loci}(t) \rangle_{T_{cell}} = 2.5$ . The production rate *per locus* in ColFRI would then be  $p = \frac{\langle R \rangle_{exp} d}{N_{loci}} = 7.7 \cdot 10^{-4} s^{-1} = \frac{1}{22} min^{-1}$ . This production rate estimate should be interpreted as an average over the cell cycle. This value is also dependent on the assumption that we observe all the RNA in the system (subsection 3.2.1). If some exon RNA remains undetected, this value is an underestimation for the *FLC* production rate. However, it is unlikely that this underestimation exceeds the error arising from the simplifications assumed in our analysis with respect to the cell cycle, i.e. up to a factor of 2. Note that this estimate is based on a cell cycle averaged *FLC* copy number and the assumption that RNA levels are in steady state. The value is not significantly different from our previous estimate of once per 17min, which assumed a constant production rate per locus throughout the cell cycle with a varying *FLC* copy number and without assuming steady state conditions (subsection 3.2.3). Our current estimate should also be interpreted as the transcription rate for a cell of average volume because we found in subsection 3.2.4 a scaling with volume. So taking into account our observed variation in RNA levels, we can estimate that the *FLC* production rate per locus ranges from once every 2.5h in the smallest cells to once per 7min in the largest cells.

To relate sense with antisense transcription, we performed smFISH

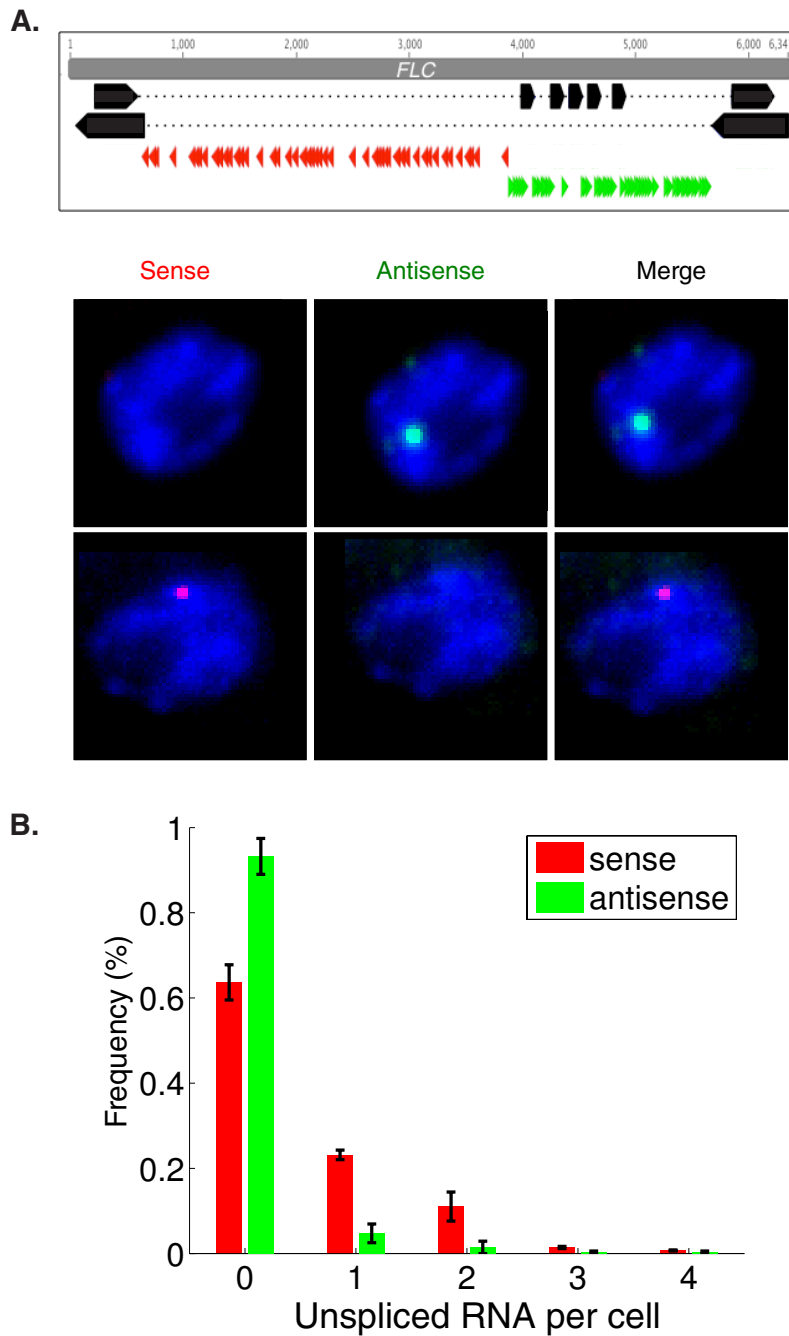


Figure 3.6: (A) Schematic of *FLC* gene with different probe sets for labelling sense unspliced RNA (red) and antisense unspliced RNA (green). Fluorescence localization of unspliced sense RNA (red), unspliced antisense RNA (green) and a merge containing unspliced sense and antisense RNA and DAPI stain (blue) in representative *Arabidopsis* root cells. (B) Frequency distributions of cellular unspliced sense RNA counts (red) and antisense (green). Values are mean  $\pm$  s.e.m. from 2 independent samples, with respectively  $N_{cells} = 118$  and 173. Images in (A) are a courtesy of Stefanie Rosa and Susan Duncan who performed the experiments.

in ColFRI with one probe set covering sense intron1 (as above) and one covering the distal antisense intron (Figure 3.6A). By measuring the number of unspliced sense  $U_s$  and antisense  $U_{as}$  spots (Figure 3.6B), we found cellular averages of  $\langle U_s \rangle = 0.54 \pm 0.05$  and  $\langle U_{as} \rangle = 0.11 \pm 0.03$ . Similar to above, the cellular unspliced RNA levels  $U$  are, in steady state, a ratio of production rate to the intron1 splicing rate:  $\langle U \rangle = \frac{pN_{loci}}{k_s}$ . In this analysis, we assume one unspliced spot corresponds to one unspliced molecule, which seems likely from our intensity analysis (subsection 3.2.5). By using the above sense production rate  $p$  this then results in an estimate for the sense intron1 splicing rate  $k_s = 3.5 \cdot 10^{-3} s^{-1} \approx \frac{1}{5} min^{-1}$ . Interestingly, this only deviates  $\sim 3x$  from the literature value we used in chapter 2 and is another line of evidence that splicing occurs co-transcriptionally. When using this intron1 splicing rate as a proxy for distal antisense splicing, we can utilize the above formula to estimate the distal antisense production rate:  $p_{distal} = 1.5 \cdot 10^{-4} s^{-1} \approx \frac{1}{2} h^{-1}$ . Below we will make use of these estimates to predict mechanistic features of *FLC* transcriptional regulation.

### 3.2.7 Estimation of Pol II elongation rates

In the following sections we combine our results from the quantitative smFISH analysis in ColFRI with those from the previous chapter that utilized population average measurements with the aim to further quantitatively characterize the transcriptional regulation of *FLC* by the autonomous pathway and *FRI*. We first use the analytic coordination model described in subsection 2.2.6 together with the parameter estimates from the previous section to infer absolute Pol II elongation rates in Col and the autonomous pathway mutant *fca-9*. That model does not specifically include cell volume, and thus resembles transcription at a locus in a cell of mean volume with average copy number. We now revisit the analysis in subsection 2.2.4 where we used nonlinear regression to fit Equation 2.11 to the *fca-9* chromatin-bound RNA fold changes to directly estimate the elongation rate changes between *fca-9* and Col. In that analysis, we found that  $v_{fold} := \frac{v^{fca-9}}{v^{Col}} = 9.8 \pm 3.8$  (mean  $\pm$  s.e.m.,  $p = 2.6 \cdot 10^{-2}$ ) and we also estimated the ratio of the absolute elongation rate in Col to the splicing rate  $\frac{v^{Col}}{k_s} = 550 \pm 140$  bp ( $p = 2.1 \cdot 10^{-3}$ ). Now we insert

our splicing rate estimate ( $k_s = 3.5 \cdot 10^{-3} \text{s}^{-1}$ ) that we experimentally observed through our smFISH measurements in ColFRI (subsection 3.2.6). It is justified to use the same value for the Col splicing rate, because the splicing efficiency seems unaffected between *Arabidopsis* accessions with and without an active *FRI* allele (data not shown). By doing this we estimate the elongation rate in Col to be  $v^{Col} = 0.12 \pm 0.03 \text{ kb/min}$ . Combining this with the  $v^{fold}$  estimates,  $v^{fca-9}$  can then range between  $0.5 \text{ kb/min}$  and  $2.0 \text{ kb/min}$  based on the chromatin-bound RNA data. The Pol II ChIP fitting procedure predicted  $v^{fold} = 8 - 12x$ , leading to a range of  $0.7 \text{ kb/min}$  to  $1.8 \text{ kb/min}$  for  $v^{fca-9}$ , which agrees well.

### 3.2.8 *FRI-mediated coordination of transcription initiation and elongation*

We next proceed by investigating quantitative transcriptional regulation of *FLC* by FRIGIDA. First, we fit our analytical coordination model that we developed for the autonomous pathway to the *FLC* mRNA fold changes (Figure 3.7A) and Pol II ChIP fold changes (Figure 3.7B, published in [256]) between ColFRI and Col, in order to determine their respective *FLC* transcription rates. A transcription-based model is justified because *FRI* is a transcriptional activator [111, 112, 257].

The methodology is the same as described in subsection 2.2.4 unless stated otherwise. Recall that this analytical model does not take into account transcriptional interference, an assumption that we will revisit in the next section. Instead, we here set the sense initiation rate equal to the respective production rate:  $F_s^{FRI} = p_s^{FRI} = 7.7 \cdot 10^{-4} \text{s}^{-1}$ . Using the observed *FLC* mRNA fold change in Figure 3.7A of  $\sim 30x$ , we estimate the sense initiation rate to be  $F_s^{Col} \approx \frac{1}{11} h^{-1}$ .

The analytical model has the following remaining *a priori* unknown parameters: the antisense initiation rates  $F_{as}^{Col}$ ,  $F_{as}^{FRI}$  and the ColFRI elongation rate  $v^{FRI}$ . The model unspliced distal RNA levels equal:

$$U_{as}^{FRI} := \frac{N_{loci} p_{as}^{FRI}}{k_s} = \frac{N_{loci} F_{as}^{FRI}}{k_s} \exp\left(-\frac{k_{pA,prox}}{v^{FRI}} L_{PII}\right).$$



The proximal and distal fold changes are equal to Equation 2.7 and Equation 2.6 respectively with the label *fca-9* replaced by *FRI*, and likewise for the model Pol II levels from subsection 2.2.4. We performed a parameter sweep to determine the unknown parameters. In here, we fix the value of  $k_{pA,s} = k_{pA,prox} = k_{pA,distal} = \frac{1}{50} s^{-1}$  as found in [41]. Furthermore, as estimated above, we set  $v^{Col} = 0.1 \text{ kb/min}$ ,  $k_s = 3.5 \cdot 10^{-3} s^{-1}$ ,  $F_s^{FRI} = 7.7 \cdot 10^{-4} s^{-1}$ ,  $F_s^{Col} = 2.9 \cdot 10^{-5} s^{-1}$ . The following criteria all needed to be met for a set of parameters to be considered as being able to fit the data:

- $\log(\mathbb{F}(distal))$  within  $\text{mean} \pm \text{s.e.m.}$  of experimental log-fold change of distal *COOLAIR* Figure 3.7A.
- $\log(0.5) \leq \log(\mathbb{F}(proximal)) \leq \text{mean} + \text{s.e.m.}$  of experimental log-fold change of proximal *COOLAIR* Figure 3.7A.
- $U_{as}^{FRI}$  within  $\text{mean} \pm \text{s.e.m.}$  of experimental values (subsection 3.2.6).
- No 5' Pol II peak observable in ColFRI despite frequent distal antisense termination, using convoluted Pol II levels:  

$$\frac{C^{FRI}(x=0)}{C^{Col}(x=100)} \leq 1.1.$$
- For comparison of gene body Pol II upregulation, using convoluted Pol II levels:  

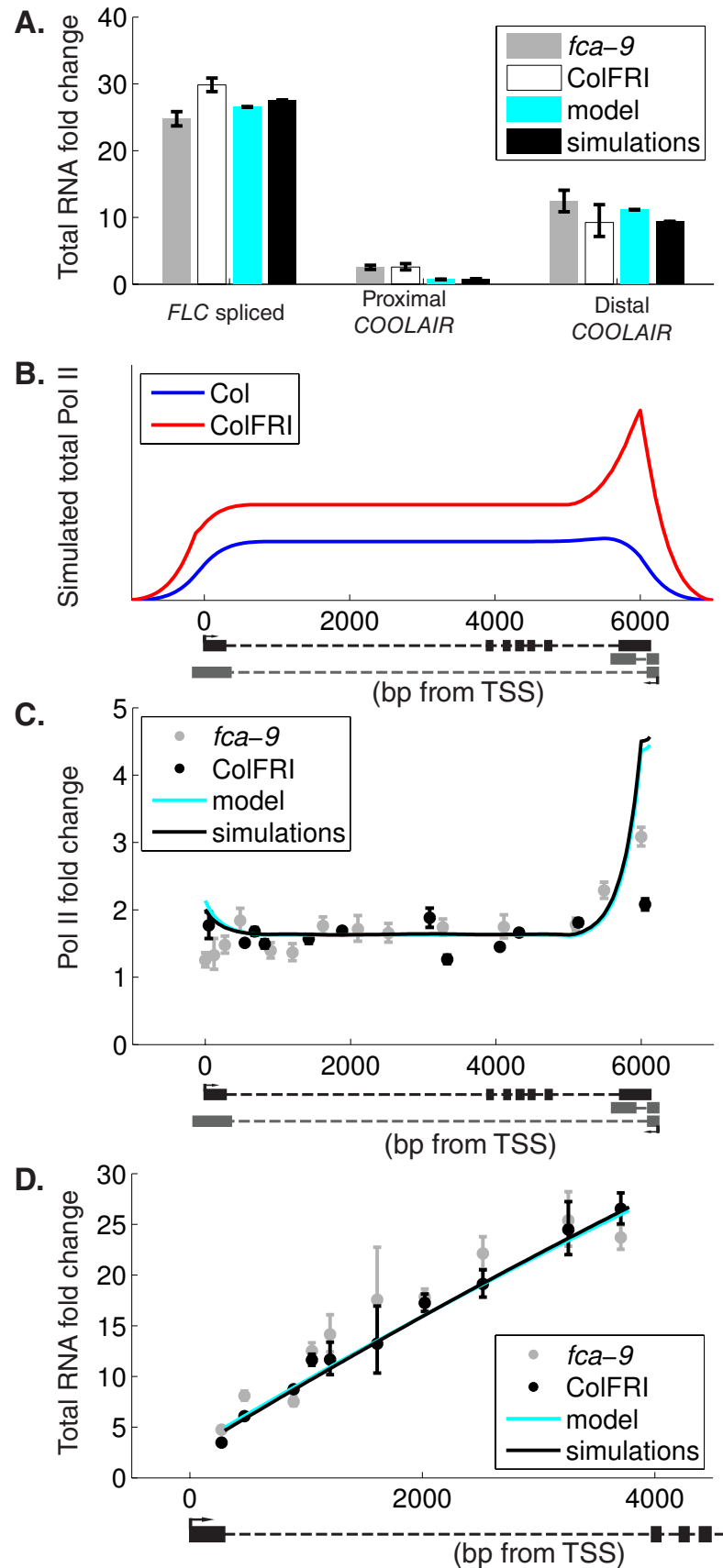
$$\log\left(\frac{C^{FRI}(x=23)}{C^{Col}(x=23)}\right) \text{ within } \text{mean} \pm \text{s.e.m. of experimental Pol II log-fold change at corresponding primer } (x = 23 \times 30 = 690 \text{ bp from TSS}) \text{ in Figure 3.7C.}$$
- Limited 3' Pol II fold change, where, using convoluted Pol II levels:  

$$\text{mean} - \text{s.e.m. of experimental log-fold upregulation at corresponding primer} \leq \log\left(\frac{C^{FRI}(x=204)}{C^{Col}(x=204)}\right) \leq \log(4.5).$$

Using this procedure, we found that the parameters values within the following ranges could fit our data.

$$\begin{aligned}
 F_{as}^{Col} &= 2.3 - 3.2 \cdot 10^{-5} s^{-1} \\
 F_{as}^{FRI} &= 1.4 - 1.9 \cdot 10^{-4} s^{-1} \\
 v^{FRI} &= 1.1 - 1.3 \text{ kb/min, leading to:} \\
 v_{fold} &= 11 - 13 \\
 \frac{F_{as}^{FRI}}{F_{as}^{Col}} &= 6 - 8
 \end{aligned}$$

Figure 3.7: (A) Total RNA fold upregulation in *fca-9* mutant and ColFRI compared to Col. Values are mean $\pm$ s.e.m. from 3 to 6 independent samples. The analytic model and stochastic simulations values are fits to the experimental data. (B) Total Pol II levels in Col and *fca-9* across *FLC* from the analytic model convolved with experimental Pol II ChIP fragment size distribution. Stochastic simulations values look similar (data not shown). (C) Experimental Pol II fold upregulation in *fca-9* mutant and ColFRI compared to Col. Experimental data for *fca-9* as in Figure 2.4D, ColFRI data as published in [256]. Analytic model and stochastic simulation fold changes are ratio of profiles shown in (B). See subsection 3.2.8 and subsection 3.2.9 for detailed parameter fitting procedures. (D) *FLC* intron1 total RNA fold change, according to analytical model and stochastic simulations, as well as the experimentally observed profiles for *fca-9* mutant and ColFRI (for *fca-9* as in Figure 2.6 ). Experiments performed by Zhe Wu, Peijin Li.



The values used for the fits in Figure 3.7 are:  $F_{as}^{Col} = 2.3 \cdot 10^{-5} s^{-1}$ ,  $F_{as}^{FRI} = 1.8 \cdot 10^{-4} s^{-1}$ ,  $v^{FRI} = 1.3 kb/min$ . This results in  $\mathbb{F}(proximal) = 0.7$ ,  $\mathbb{F}(distal) = 11$  and  $U_{as}^{FRI} = 0.13$ , consistent with the experimental values (Figure 3.7A and subsection 3.2.6). This shows that the antisense initiation rates are relatively constrained by this fitting procedure.

The Pol II ChIP fit (Figure 3.7B,C) predicts that elongation rate changes also occur between Col and ColFRI with an estimated elongation rate in ColFRI of  $v^{FRI} = 1.3 kb/min$ . It must be noted that the ColFRI and *fca-9* Pol II fold changes differ in some aspects (Figure 3.7C), with the *FRI* assays consisting of fewer replicates [256]. We cannot rule out that some of these differences are technical in nature. Using the above parameters, we then also predicted the total intronic RNA profile (Figure 3.7D) employing only the lariat degradation rate  $k_l = 1.5 bp/s$ , with the same value as for the *fca-9* and *fld-4* analysis (subsection 2.2.6). The intronic fold changes provide further evidence that elongation rate changes, as predicted by the Pol II ChIP fitting procedure, also occur between Col and ColFRI.

In this section we estimated absolute (mean) Pol II initiation and elongation rates in Col and ColFRI from a combination of smFISH and population average measurements. We conclude that it is highly likely that *FRI* mediates coordinated changes in transcription initiation and elongation, as observed between Col and autonomous pathway mutants.

### 3.2.9 *Stochastic coordination model predicts that fast Pol II elongation enables a high FLC expression state*

Our absolute quantitative rate estimates enabled us to investigate features of *FLC* regulation that would otherwise remain unclear. Here we describe the potential involvement of transcriptional interference and the functionality of elongation rate changes mediated by *FRIGIDA*.

We performed stochastic simulations representing transcription at one *FLC* locus (using a Gillespie algorithm [218] in C++) in a low (Col) and high expression *FLC* state (ColFRI). Our simulations did

not specifically include cell volume, and thus resemble transcription at one locus in a cell of mean volume with average copy number. The reactions and reactants are identical to the model described in subsection 2.2.10 with the exception that we did not simulate FLD, A mark and unbound Pol II dynamics. Instead, initiation and elongation occurred with constant rates as described below. All reactions and reactants in this model are illustrated in Figure 3.8. All parameters are either constrained by the data described in the previous sections or taken to be literature values (Table 3.3).

In case of an unoccupied sTSS, the propensity is:

$$p_t = F_s.$$

Similar rules apply for antisense with initiation rate  $F_{as}$  at the asTSS. The propensity for a sense (antisense) Pol II at site  $i$ ,  $P_i$ , to elongate to an unoccupied site  $i + 1$  ( $i - 1$ ) is given by:

$$p_t = k_{elon}P_i.$$

We also modelled *FLC* mRNA (species *sFLC* in the model) degradation as a first order reaction with rate  $d$ , a first order splicing reaction with rate  $k_s$  for full length distal RNA (*unasFLC<sub>distal</sub>*) at the locus to result in spliced distal *COOLAIR* (*asFLC<sub>distal</sub>*) (Figure 3.8).

The amount of simulations and model output methodology were as described in subsection 2.2.10, with the addition that we would like to compare our model with the experimental values for the number of unspliced sense  $U_s$  and antisense  $U_{as}$  spots per cell (Figure 3.6B), as detailed below. Therefore from the model, we also output the amount of unspliced sense intron1 RNA residing at the locus. Sense intron1 RNA is a sum of full length (intron1 containing) RNA and Pol II bound RNA:  $U_s := unsFLC + s2FLC + \sum_{i=I1A}^{L-1} sP_i + sS2_i$ . We refer to subsection 2.2.10 for a description of the reactants mentioned here. We also output the amount of unspliced distal antisense RNA residing at the locus, as a sum of full length unspliced RNA and antisense Pol II that has transcribed beyond site *I1A*:  $U_{as} := unasFLC_{distal} + \sum_{i=0}^{I1A} aP_i$ . The site *I1A* is chosen because all unspliced antisense smFISH probes are located more upstream towards the antisense TSS (Figure 3.1). Therefore elongating Pol II that have transcribed this

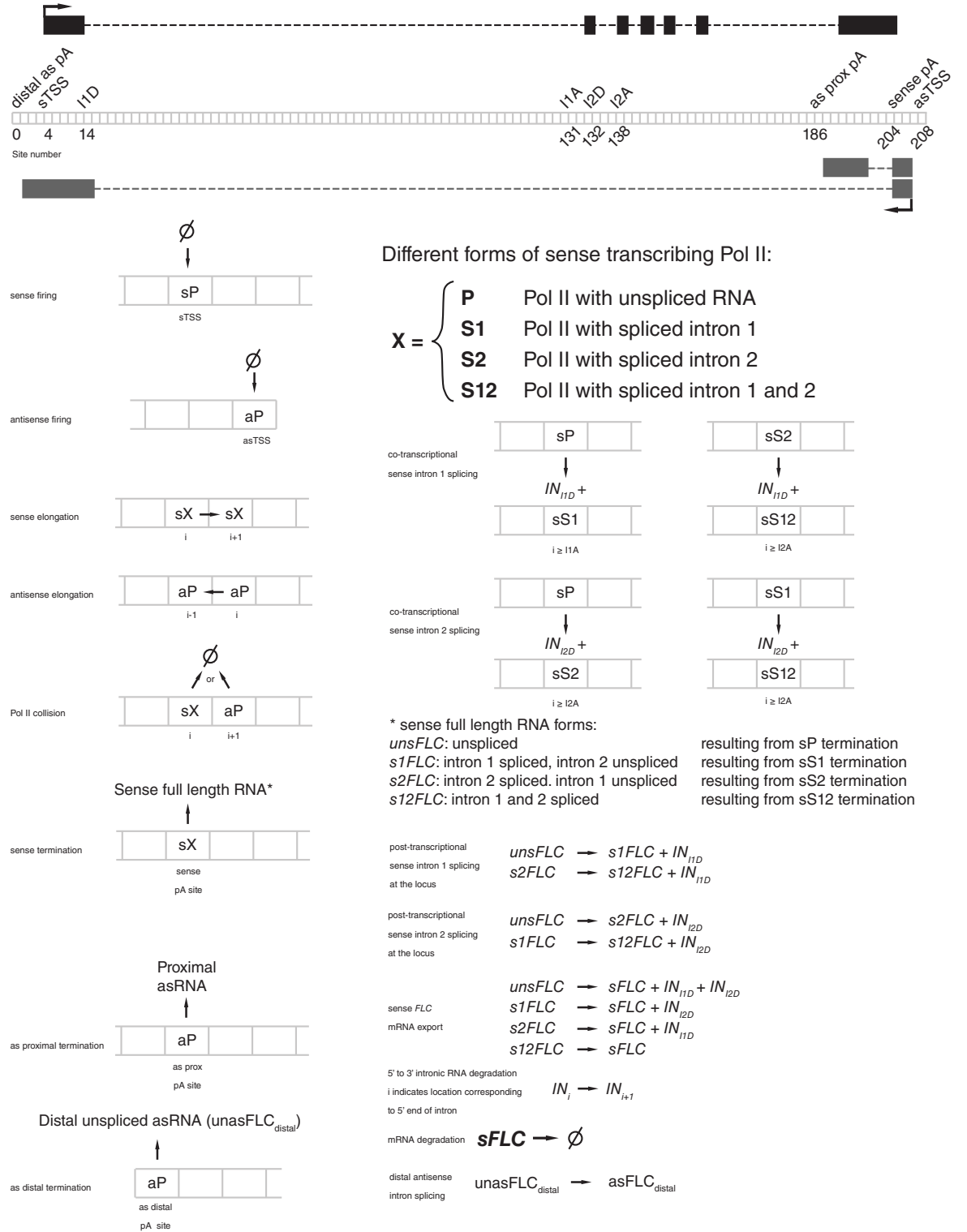


Figure 3.8: Schematic of the processes taken into account in the stochastic simulations of *FLC* transcription in Col and ColFRI, with coordinated initiation and elongation rates. Black (grey) rectangular boxes in the schematic indicate sense (antisense) exons.

Parameter	Description	Value	Notes
$L_{PII}$	Length of a Pol II site along the <i>FLC</i> gene	$30bp$	This length corresponds to the size of a Pol II footprint [219].
$L$	Length of <i>FLC</i> gene in sites	209	For simulation purposes <i>FLC</i> is divided into $L$ sites of length $L_{PII}$ .  At most one Pol II can reside at each site.
$F_s$	Sense initiation rate	$2.9 \cdot 10^{-5} s^{-1}_{Col}$ , $7.8 \cdot 10^{-4} s^{-1}_{(ColFRI)}$ , $7.8 \cdot 10^{-4} s^{-1}_{(ColFRI+slow)}$	Constrained by experiment subsection 3.2.8.
$F_{as}$	Antisense initiation rate	$2.3 \cdot 10^{-5} s^{-1}_{Col}$ , $1.8 \cdot 10^{-4} s^{-1}_{(ColFRI)}$ , $1.8 \cdot 10^{-4} s^{-1}_{(ColFRI+slow)}$	Constrained by experiment subsection 3.2.8.
$k_{elon}$	Elongation rate	$0.056 s^{-1}_{Col}$ , $0.73 s^{-1}_{(ColFRI)}$ , $0.056 s^{-1}_{(ColFRI+slow)}$	Constrained by experiment, subsection 3.2.7  and subsection 3.2.8 .
$k_s$	Splicing (cleavage) rate	$0.0035 s^{-1}$	Constrained by experiment subsection 3.2.6
$k_l$	Intronic RNA degradation rate	$0.05 s^{-1}$	Constrained by experiment [44], 5' to 3' degradation assumed in line with [214], corresponds to $1.5bps^{-1}$ .
$k_{pA}$	Pol II termination rate at poly(A) site	$0.02 s^{-1}$	Constrained by experiment [41].
$k_{col}$	Pol II termination rate due to sense/ antisense Pol II collision	$0.02 s^{-1}$	Taken to be the same value as $k_{pA}$ in absence of further information.
$k_x$	mRNA export rate	$1.7 \cdot 10^{-3} s^{-1}$	Constrained by experiment [41].
$d$	degradation rate	$3.25 \cdot 10^{-5} s^{-1}$	Constrained by experiment subsection 3.2.2.

Table 3.3: Parameter values used in the stochastic coordination model in ColFRI. (Col) and (ColFRI) indicate parameter values that are specific for simulations of those genotypes. 'ColFRI+slow' indicates the simulations of a high *FLC* expression, as in ColFRI, in the presence of a slow Pol II elongation rate.

upstream part could in principle be detected.

In the simulations (Figure 3.7, Figure 3.9), we set the initiation and elongation rates the same as for the corresponding analytical model fits. In this way we could test whether effects that were not taken into account in our analytical effects, for instance transcriptional interference, would affect *FLC* regulation. In the simulations, one of the two Pol II terminates as a consequence of a sense/antisense Pol II collision, leading to a lower overall sense and distal antisense production rate. The model mean cellular *FLC* levels,  $\langle FLC \rangle_{sim} = N_{loci} \langle sFLC \rangle =$

57 (with  $\langle sFLC \rangle$  representing the average over 1000 simulation values) resemble the experimental values  $59 \pm 2$  (subsection 3.2.1) and analytical model values  $\langle FLC \rangle_{model} = N_{loci} \frac{F_{as}^{FRI}}{d} = 59$  well. This indicates that transcriptional interference plays a marginal role in *FLC* regulation in ColFRI, most likely due to the relatively low absolute *COOLAIR* initiation rate and the relatively fast elongation rates.

Our mean cellular unspliced sense RNA levels  $\langle U_s^{FRI} \rangle_{sim} = N_{loci} \langle U_s^{FRI} \rangle = 0.46$  resembled the analytical model values  $\langle FLC \rangle_{model} = N_{loci} \frac{F_{as}^{FRI}}{k_s} = 0.55$  reasonably well. They were also similar to the experimentally determined range of the mean cellular sense unspliced spot count  $\langle U_s \rangle_{exp} = 0.54 \pm 0.05$  (subsection 3.2.6). Under the assumption that one spot resembles one RNA, these values are comparable. We found a similarly good correspondence for unspliced antisense RNA levels:  $\langle U_{as}^{FRI} \rangle_{sim} = N_{loci} \langle U_{as}^{FRI} \rangle = 0.11$ ,  $\langle U_{as} \rangle_{model} = N_{loci} \frac{F_{as}^{FRI}}{k_s} = 0.13$  and  $\langle U_{as} \rangle_{exp} = 0.11 \pm 0.03$ . We could also fit all other previously described population average data with analytical model and stochastic simulations, which resembled each other well (Figure 3.7).

Given that our Gillespie algorithm models initiation, elongation and degradation as a sequence of first order reactions, we expect Poisson distributions for the various *FLC* forms (subsection 3.4.1). Indeed simulation *FLC* mRNA levels are well approximated by a Poisson distribution with mean  $\langle sFLC \rangle = 23$ , a mean very similar to  $\frac{F_s^{FRI}}{d} = 24$  (Figure 3.9A,B). This is in accordance with our experimental observations that the volume corrected noise levels are close to the Poisson limit. Furthermore, the simulations resemble well the expected Poisson distributions of unspliced RNA (with mean  $\frac{F_s^{FRI}}{k_s} = 0.2$ ) and sense exon RNA (with mean  $\frac{F_s^{FRI}}{k_x} = 0.5$ ) at the locus (Figure 3.9B,C) with the latter based on a literature mRNA export value of  $\frac{1}{10} min^{-1}$ . Consistent with our quantitative image analysis (subsection 3.2.5) we find that accumulation of multiple RNAs only occurs in a small fraction of cells (Figure 3.9B,C).

Lastly, we tested whether the elongation rate increase in the high expression state had any functional consequence on the *FLC* production rate, by simulating a scenario with initiation rates as in ColFRI,

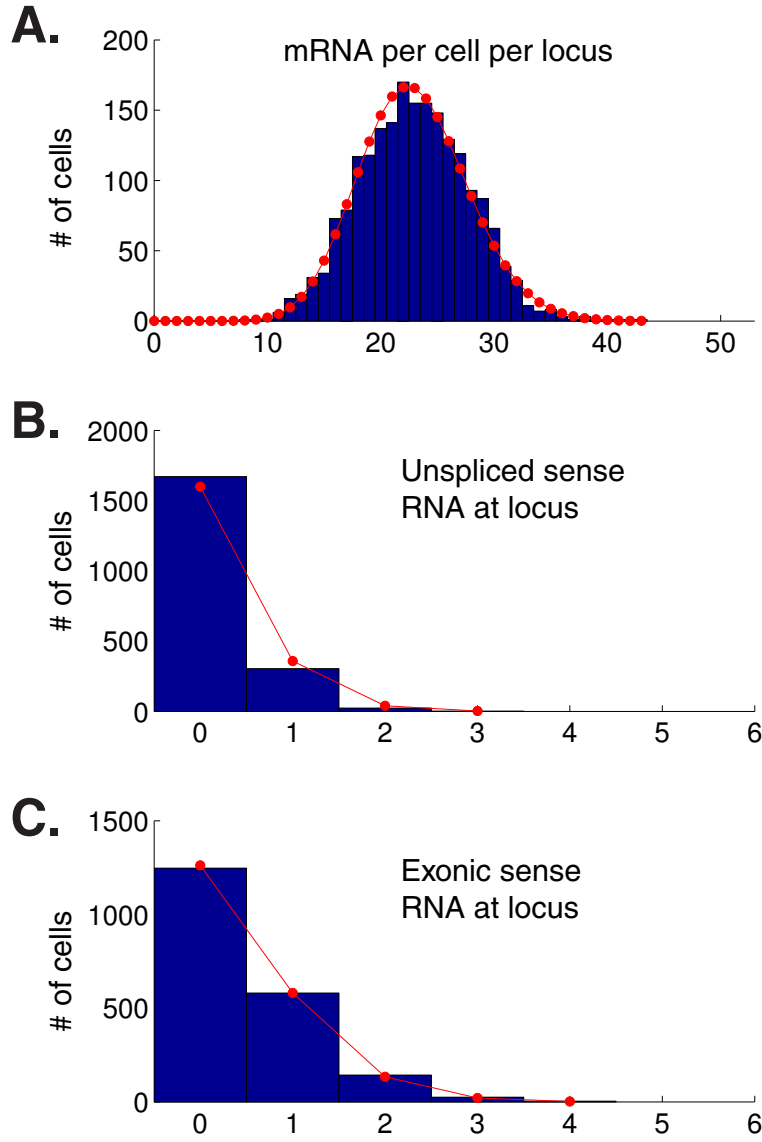


Figure 3.9: (A) Histograms of cellular exon RNA per locus predicted by ColFRI stochastic simulations (blue) and a Poisson distribution fit (red) with mean  $\langle sFLC \rangle = 23$ , a value very similar to  $\frac{F_s^{FRI}}{d} = 24$ . (B) As in (A) for unspliced RNA at a locus, Poisson distribution mean  $\frac{F_s^{FRI}}{k_s} = 0.2$ . (C) As in (A) for exon RNA at a locus, Poisson distribution mean  $\frac{F_s^{FRI}}{k_x} = 0.5$ .



but an elongation rate as in Col ( $0.1\text{kb}/\text{min}$ ). The resulting mean mRNA levels  $\langle sFLC \rangle$  dropped  $\sim 15\%$  compared to the ColFRI levels. This suggests that fast elongation rates could be functionally important to prevent transcriptional interference and thereby enable a high *FLC* expression state.

### 3.3 Discussion

Gene expression is a fundamentally stochastic process, with randomness in transcription and translation leading to significant cell-to-cell variations in mRNA and protein levels [234]. Here, we have quantitatively characterized the variation in *Arabidopsis FLC* expression throughout the complete RNA life cycle. We find that cellular *FLC* transcription variation scales predominantly with cell volume, a feature that has recently been observed for several genes in mammalian cells [235]. Furthermore we found that, when corrected for volume, *FLC* transcription, processing and degradation are well described by Poisson processes. This is relevant, because it indicates that these stochastic events occur in an uncorrelated manner, in contrast to regulation by transcriptional bursts [222], which are observed in organisms ranging from bacteria and yeast to mammals [238–240]. We characterized for the first time the kinetics of transcription events in plants, which seem, at least for *FLC*, to be more similar to certain transcriptionally active loci in yeast that also exhibit uncorrelated initiation events [258]. The Poisson characteristic also suggests that for a given volume, the expression variation is dominated by intrinsic noise, with little influence from any other factors (besides volume) which could in principle contribute to extrinsic variation. Lastly, our results do not favour cellular bistability in non-vernalized conditions, contrasting with the epigenetic regulation during and after cold [138, 139, 259].

Our findings significantly advance our understanding of expression variation in plants, which possess intricate regulatory mechanisms to ensure proper development and adequate responses to environmental stimuli [90]. *FLC* gene expression quantitatively correlates with flowering time both in various natural accessions and mutants [89, 260–266]. We propose that transcriptional regulation could have

evolved to generate quantitatively fine-tuned *FLC* expression with relatively little cellular concentration variation in a given genotype. Such fine-tuned *FLC* expression could then ensure a precisely timed floral transition. Consistently, the *FLC* half life is one of the longest among *Arabidopsis* transcription factors [248], potentially to buffer against fluctuations that are associated with faster kinetics [267]. A large volume-independent variation would leave certain cells with a low *FLC* concentration. In those cells, that could result in an untimely induction of a floral activator such as *FT*. The *FT* protein is known to be mobile [231–233], so that such aberrant *FT* induction in those cells could have concomitant downstream consequences on the timing of flowering.

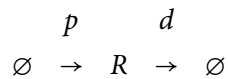
Our findings also contribute to our mechanistic understanding of *FLC* regulation. *COOLAIR* 3' processing can occur at a proximal poly(A) site and a distal poly(A) site as reviewed in chapter 1. In the stochastic model developed in subsection 2.2.10, FCA-mediated proximal processing results in FLD action to repress *FLC* and *COOLAIR* transcription initiation. Distal processing did not directly affect transcriptional regulation. It has been proposed that the act of antisense transcription to the distal site could result in activation of sense initiation [97]. We concluded that this mechanism is not sufficient to explain the data (subsection 2.2.10). However, we do not rule out that readthrough or distal *COOLAIR* processing could activate *FLC* transcription [256]. In yeast, a genetic toggle involving lncRNAs leads to variegated gene expression (see chapter 1 for details). This has raised the hypothesis that a toggle between proximal and distal poly(A) sites could also lead to varied *FLC* expression [127]. In contrast, we find here that *FLC* transcription seems tightly controlled rather than variegated. We conclude that a poly(A) site toggling mechanism, if at all present, contributes little to *FLC* expression variation. Moreover, our stochastic coordination model did generate *FLC* distributions that are well approximated by a Poisson distribution (Figure 3.9). Therefore we still favour a mechanism where FCA-mediated proximal *COOLAIR* processing leads to chromatin regulation.

Global transcriptional activity scaling with cell volume is most likely caused by Pol II abundance [235]. It remains an interesting question what then determines the gene-specific transcription rate per unit of volume. Certain histone modifications, e.g. H3K4me2/3, have been proposed to provide a short-term memory for transcription events [20]. At *FLC*, for an average cell volume, we have found the mean production rate (per locus) in ColFRI to be around once per  $\sim 22\text{min}$  and that elongation in this state would typically take several minutes. So there is often one, but most likely not multiple Pol II engaged in *FLC* transcription at a given time. Given the turnover time scales of histones [220] and its specific modifications [268, 269] histone modifications such as H3K4me2/3 and H3K36me3 remain interesting candidates to determine the transcription rate. Activity of autonomous pathway components such as FLD could result in removal of these histone modifications [120], whilst FRI induces active placement of these marks [111, 112] resulting in a quantitative balance in non-vernalized conditions. Potentially the chromatin state could affect elongation through a Pol II gearing mechanism where the chromatin state determines the speed of elongation, either directly or through recruitment of elongation factors. Fast elongation could then enable fast (re)-initiation. Consistently, our modelling suggests fast elongation helps to minimize transcriptional interference and enable a fast RNA production rate.

### 3.4 Materials and Methods

#### 3.4.1 *Probability distributions arising from birth-death processes*

In this section we investigate the probability distributions when RNA production and degradation are modelled by birth-death processes. We start with a one stage process  $\{R(t)\}_{t \geq 0}$  with constant transition probability per unit time for production ( $p$ ) and degradation ( $d$ ) [244]:



Thus  $R(t)$  indicates the (exon) RNA levels at time  $t$  and  $p$  and  $d$  resemble the mean RNA production and degradation rates that would

occur in the analogous deterministic system described by

$$\frac{dR(t)}{dt} = p - dR(t).$$

Let  $\mathbb{P}(r, t) := \mathbb{P}(R(t) = r)$ , the probability of having  $R(t) = r$  exon RNAs at time  $t$ . Then the master equation for  $\mathbb{P}(r, t)$  is given by [244]:

$$\frac{d\mathbb{P}(r, t)}{dt} = p\mathbb{P}(r-1, t) - p\mathbb{P}(r, t) + d(r+1)\mathbb{P}(r+1, t) - dr\mathbb{P}(r, t). \quad (3.4)$$

In steady state ( $\frac{d\mathbb{P}(r, t)}{dt} = 0$ ), the stationary solution  $\mathbb{P}_s(r)$  is straightforwardly obtained by solving Equation 3.4 recursively [244]:

$$\mathbb{P}_s(r) = \mathbb{P}_s(0) \prod_{b=1}^r \frac{p}{db} = \mathbb{P}_s(0) \frac{\left(\frac{p}{d}\right)^r}{r!}.$$

Because  $\{\mathbb{P}_s(r)\}_{r \geq 0}$  is a probability measure, its normalization determines  $\mathbb{P}_s(0)$  so that we obtain  $\mathbb{P}_s(r) = \frac{\left(\frac{p}{d}\right)^r}{r!} \exp\left(-\frac{p}{d}\right)$ , a Poisson distribution with mean  $\frac{p}{d}$ . More generally, the probability measure  $\mathbb{P}(r, t)$  is a (time-dependent) Poisson distribution with mean  $\frac{p}{d}(1 - \exp(-dt))$  in the case where there is no initial exon RNA present ( $\mathbb{P}(r, 0) := \delta_{0,r}$ ), as derived in [245] using the method of characteristics (see below for more details). We conclude that a one stage birth-death process generates a Poisson distribution.

To model RNA levels after transcription inhibition, we consider a pure death process by setting the production rate  $p$  to zero in Equation 3.4:

$$\begin{array}{c} d \\ R \rightarrow \emptyset \end{array}$$

$$\frac{d\mathbb{P}(r, t)}{dt} = d(r+1)\mathbb{P}(r+1, t) - dr\mathbb{P}(r, t). \quad (3.5)$$

To derive the probability distribution  $\mathbb{P}(r, t)$  given any distribution  $\{\mathbb{P}(r, 0)\}_{r \geq 0}$  as an initial condition we define the generating function  $G(w, t) := \sum_{b=0}^{\infty} w^b \mathbb{P}(b, t)$  [244]. We then convert Equation 3.5 into a partial differential equation (PDE) for  $G(w, t)$ :

$$\frac{\partial G(w, t)}{\partial t} = d \frac{\partial G(w, t)}{\partial w} - dw \frac{\partial G(w, t)}{\partial w}. \quad (3.6)$$

Using the methods of characteristics [270], we can solve Equation 3.6 analytically for  $G(w, t)$ , given the Dirichlet boundary conditions  $G(w, 0) =$

$\sum_{b=0}^{\infty} w^b \mathbb{P}(b, 0)$ . The aim is to find the solution  $(w, t, G(w, t))$ , resembling a surface  $S$  in  $\mathbb{R}^3$ . To construct  $S$ , let us consider a curve  $C$  on  $S$ :

$$C : [0, \infty] \rightarrow \mathbb{R}^3, \quad q \mapsto (w(q), t(q), G(w(q), t(q))).$$

Recall from calculus that the tangent to  $C$  is given by  $\left(\frac{dw(q)}{dq}, \frac{dt(q)}{dq}, \frac{dG(w(q), t(q))}{dq}\right)$  and the normal to  $C$  equals  $\left(\frac{\partial G(w, t)}{\partial w}, \frac{\partial G(w, t)}{\partial t}, -1\right)$ . Because these vectors are orthogonal at each point along the curve, we find that:

$$\frac{dw(q)}{dq} \frac{\partial G(w, t)}{\partial w} + \frac{dt(q)}{dq} \frac{\partial G(w, t)}{\partial t} - \frac{dG(q)}{dq} = 0.$$

By equating the above equation to Equation 3.6 we find:

$$\begin{aligned} \frac{dw(q)}{dq} &= d(w - 1) \\ \frac{dt(q)}{dq} &= 1 \\ \frac{dG(q)}{dq} &= 0. \end{aligned}$$

This equation can straightforwardly be solved analytically. Now we impose the boundary conditions:  $t(q = 0) = 0$ ,  $w(q = 0) = w_0$ , then we have:

$$\begin{aligned} w(q) &= 1 - \exp(dq) + w_0 \exp(dq) \\ t(q) &= q \\ G &= \sum_{b=0}^{\infty} w_0^b \mathbb{P}(b, 0). \end{aligned} \tag{3.7}$$

Combining the expressions leads to the desired expression for  $G(w, t)$ :

$$G(w, t) = \sum_{b=0}^{\infty} (w \exp(-dt) + 1 - \exp(-dt))^b \mathbb{P}(b, 0).$$

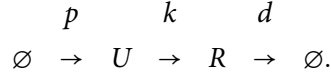
Now by the definition of the generating function we can make use of its Taylor series expansion to obtain  $\mathbb{P}(r, t)$ :

$$\mathbb{P}(r, t) = \frac{1}{r!} \frac{\partial^r}{\partial w^r} G(w, t) \Big|_{w=0}.$$

Let  $M := \max\{r : \mathbb{P}(r, 0) > 0\}$ , finite in our case. Then we obtain the desired result:

$$\mathbb{P}(r, t) = \sum_{b=r}^M \binom{b}{r} (\exp(-dt))^r (1 - \exp(-dt))^{b-r} \mathbb{P}(b, 0)$$

Lastly we have investigated a two-stage birth-death process:



The analogous deterministic system is described by:

$$\begin{aligned} \frac{dU(t)}{dt} &= p - kU(t) \\ \frac{dR(t)}{dt} &= kU(t) - dR(t). \end{aligned}$$

This analysis is relevant, because it is *a priori* not clear if the expected RNA distribution would still be Poissonian if the RNA life cycle would be considered as a multi step process, a much more realistic scenario. By considering a two-stage birth-death process (Equation 3.8), we show that both RNA species, for instance resembling unspliced RNA and exon RNA are both Poisson distributed. This shows that our conclusion that the experimentally observed super-Poissonian distribution cannot be explained by a birth-death process with constant rates is not dependent on the fact that RNA production and degradation are modeled as one step processes.

We straightforwardly generalize to a joint probability distribution  $\mathbb{P}(u, r, t) := \mathbb{P}(U(t) = u, R(t) = r)$  for both exon RNA  $R(t)$  and  $U(t)$ , that we label for convenience as unspliced RNA. We refer to parameter  $k$  as the RNA processing rate. It is clear that these two variables can represent other species, for instance  $U$  could represent an exon RNA at the locus and  $k$  the export rate. The master equation is then given by [244]:

$$\begin{aligned} \frac{d\mathbb{P}(u, r, t)}{dt} &= p\mathbb{P}(u-1, r, t) - p\mathbb{P}(u, r, t) \\ &\quad + k(u+1)\mathbb{P}(u+1, r-1, t) - ku\mathbb{P}(u, r, t) \\ &\quad + d(r+1)\mathbb{P}(u, r+1, t) - dr\mathbb{P}(u, r, t). \end{aligned} \quad (3.8)$$

We state in brief here our derivation to obtain  $\mathbb{P}(u, r, t)$ , again using the method of characteristics. Let  $G(v, w, t) := \sum_{a,b=0}^{\infty} v^a w^b \mathbb{P}(a, b, t)$  be the moment generating function. Then the PDE for  $G(v, w, t)$

corresponding to Equation 3.8 is:

$$\begin{aligned}\frac{\partial G(v, w, t)}{\partial t} = & p(v-1)G(v, w, t) \\ & + k(w-v)\frac{\partial G(v, w, t)}{\partial v} \\ & + d(1-w)\frac{\partial G(v, w, t)}{\partial w}.\end{aligned}$$

Now we consider the parameterization

$$C: [0, \infty] \rightarrow \mathbb{R}^4, \quad q \mapsto (v(q), w(q), t(q), G(v(q), w(q), t(q)))$$

along a characteristic curve. Then we obtain the corresponding system of ODEs:

$$\begin{aligned}\frac{dv(q)}{dq} &= k(v-w) \\ \frac{dw(q)}{dq} &= d(w-1) \\ \frac{dt(q)}{dq} &= 1 \\ \frac{dG(q)}{dq} &= p(v-1)G.\end{aligned}$$

Solutions for  $w(q)$  and  $t(q)$  are as found in Equation 3.7. We can insert the expression for  $w(q)$  into the first order ODE for  $v(q)$  and solve it analytically to obtain:

$$v(q) = 1 - \exp(kq) + \frac{k}{d-k} [1 - w_0] [\exp(dq) - \exp(kq)] + v_0 \exp(kq).$$

The first order ODE for  $G(q)$  can be solved as:

$$G(q) = G_0 \exp\left(p \int_0^q [v(q') - 1] dq'\right).$$

Assuming the boundary condition that there are no RNAs present initially,  $\mathbb{P}(u, r, 0) = \delta_{u,0} \delta_{r,0}$ , then  $G_0 = G(v, w, t=0) = 1$ . When we insert the above expression for  $v(q)$ , integrate and then replace  $v_0$  and  $w_0$  by their respective expressions in terms of  $v$  and  $w$ , we obtain the desired expression for the generating function:

$$\begin{aligned}G(v, w, t) = & \exp\left\{\frac{p}{k}[v-1][1-\exp(-kt)]\right\} \\ & \times \exp\left\{\frac{p}{k-d}[w-1]\left[\frac{k}{d}-1+\exp(-kt)-\frac{k}{d}\exp(-dt)\right]\right\}\end{aligned}$$

Again by invoking the Taylor expansion for  $G(v, w, t)$

$$\mathbb{P}(u, r, t) = \frac{1}{u!r!} \frac{\partial^u}{\partial v^u} \frac{\partial^r}{\partial w^r} G(v, w, t) \Big|_{v,w=0},$$

we then find that the levels of  $R(t)$  and  $U(t)$  are independently distributed and both follow (time-dependent) Poisson distributions:

$$\mathbb{P}(u, r, t) = \frac{\lambda(t)^u}{u!} \exp(-\lambda(t)) \frac{\mu(t)^r}{r!} \exp(-\mu(t))$$

with means

$$\begin{aligned} \lambda(t) &= \frac{p}{k} (1 - \exp(-kt)) \\ \mu(t) &= \frac{p}{d} \left( 1 + \frac{d}{k-d} \exp(-kt) - \frac{k}{k-d} \exp(-dt) \right). \end{aligned}$$

The expression for  $\lambda(t)$  was expected given the results for a one-stage birth-death process stated above [245]. To find the steady state distributions we take the limit of  $t \rightarrow \infty$ . This results in Poisson distributions for  $U$  with mean  $\frac{p}{k}$  and for  $R$  with mean  $\frac{p}{d}$ . This procedure can in principle be generalized to a multi-step birth-death process. We conclude that with constant rates the distributions of the variables are both Poissonian.

In subsection 3.2.1 we found that the observed exon RNA distribution is super-Poissonian. That could arise either because RNA production and/or degradation rates are not constant or because the stochastic events giving rise to the exon RNA distribution are not uncorrelated. In subsection 3.2.2 we found that degradation is well described by a pure death process (Equation 3.5). Thus degradation does satisfy the above conditions. Then in subsection 3.2.4 we found that the production rate correlates with volume, so that indeed the RNA production rate appears not to be constant. This volume scaling satisfactorily explains the observed super-Poissonian variation.

#### 3.4.2 *Model with constant production rate and varying $FLC$ copy number*

Here we derive the solution to the ODE model (Equation 3.3) describing the mean cellular RNA levels  $R(t)$  over time by modelling production and degradation throughout the cell cycle as described in subsection 3.2.3. Recall that we assume that the production rate per locus  $p$  is assumed to be constant. Throughout the cell cycle the *FLC* copy number increases from 2 to 4, according to Equation 3.2, in line with experimentally observed cell cycle time scales [250, 251].



We use a constant degradation rate  $d$  (unit:  $hour^{-1}$ ) as found in subsection 3.2.2. Therefore the ODE model can be described as (with  $t$  in hours):

$$\frac{dR(t)}{dt} = \begin{cases} 2p - dR(t) & \text{if } 0 \leq t < 7, \\ 3p - dR(t) & \text{if } 7 \leq t < 8, \\ 4p - dR(t) & \text{if } 8 \leq t < 14.5 \\ -dR(t) & \text{if } 14.5 \leq t \leq 17. \end{cases}$$

Solving these first order ODEs on their domain and assuming  $R(t)$  to be continuous within the cell cycle, we obtain:

$$R(t) = \begin{cases} \frac{p}{d} [2 + \exp(-dt)(K_0 - 2)] & \text{if } 0 \leq t < 7, \\ \frac{p}{d} [3 + \exp(-d(t-7))(-1 + (K_0 - 2)\exp(-7d))] & \text{if } 7 \leq t < 8, \\ \frac{p}{d} [4 + \exp(-d(t-8))(-1 - \exp(-d) + (K_0 - 2)\exp(-8d))] & \text{if } 8 \leq t < 14.5 \\ \frac{p}{d} \exp(-d(t-14.5)) & \text{if } 14.5 \leq t \leq 17. \\ \times (4 - \exp(-6.5d) - \exp(-7.5d) + (K_0 - 2)\exp(-14.5d)). & \end{cases} \quad (3.9)$$

Here  $K_0 := \frac{R(t=0)}{\frac{p}{d}}$  is *a priori* unknown. When we assume that RNA levels halve upon cell division at  $t = T_{cell} = 17$  (hours), so that

$$2R(t=0) = 2\frac{p}{d}K_0 = R(T_{cell}),$$

we find:

$$K_0 = \frac{4\exp(-2.5d) - \exp(-9d) - \exp(-10d) - 2\exp(-dT_{cell})}{2 - \exp(-dT_{cell})}.$$

The solution  $R(t)$  is visualized in Figure 3.3A. Note that in this model steady state is reached exponentially fast (Equation 3.9). However, in our particular solution, RNA levels never reach steady state throughout the cell cycle (Figure 3.3A). This feature arises because the associated time scale to reach steady state is the relatively long *FLC* half life of 5.9h (see subsection 3.2.2). By setting the model RNA levels, averaged over the cell cycle  $\langle R(t) \rangle_{T_{cell}} = \frac{1}{T_{cell}} \int_0^{T_{cell}} R(\tau) d\tau$ , equal to the observed experimental mean RNA levels  $\langle R \rangle_{exp} = 59$  we estimate the production rate per locus to be  $p = \frac{1}{17} min^{-1}$ .



## EQUAL SPACING OF BACTERIAL PLASMIDS

---

### 4.1 Introduction

After the description of our primary research results on *FLC* regulation in the previous two chapters, we now proceed with our primary results on regular plasmid positioning in bacteria. Although the biological topics of the research have been different, the mathematical modelling techniques used have been very similar.

Low copy number plasmids in bacteria require segregation for stable inheritance through cell division. As discussed in chapter 1, this is often achieved by a *parABC* locus, comprising an ATPase ParA, DNA-binding protein ParB and a *parC* region, encoding ParB-binding sites. These components space plasmids equally over the nucleoid, yet the underlying mechanism has not been understood. In this chapter, we begin by showing mathematically that competition between dynamic ParA concentrations on either side of a plasmid can lead to equal plasmid spacing. This mechanism relies on an ability of a plasmid to move towards higher ParA concentrations, but the exact means of such movement is not important.

Several mechanisms have been proposed to explain ParA-mediated plasmid movement. We therefore also investigate which aspects of the polymer and gradient mechanisms are required and sufficient to explain the observed plasmid translocation and equal spacing over the nucleoid. We then examine whether predictions from such models are borne out experimentally. We define a computational diffusion/immobilization model where nucleoid-bound ParA-ATP can anchor diffusing plasmids. We show that diffusion/immobilization can in principle space mobile plasmids equally over the nucleoid. However, experiments measuring increased plasmid mobility in the presence of the pB171 *parABC* locus (*par2*), lead us to disfavor this model. Instead we favour a directed motion mechanism in which

ParA structure formation provides directionality to plasmid motion thereby speeding up plasmid movement. The directed motion model produces robust equal plasmid spacing with, on average, relatively symmetric ParA distributions, a prediction we also verify experimentally. Furthermore, we show experimentally that ParA organization is dependent on the underlying nucleoid structure, with nucleoid disruption resulting in perturbed plasmid positioning. Our combination of modeling and experiments has for the first time uncovered a robust mechanism for plasmid spacing that unifies previous proposals.

## 4.2 Results

### 4.2.1 *ParB-GFP foci are spaced equally over the nucleoid.*

To study *par2*-mediated plasmid segregation, we investigated ParB-GFP localization, expressed from a *par2*-carrying test plasmid (see [271] in appendix for details on strains and plasmids used). The *par2* locus containing the *parB::sfGFP* fusion is fully functional as judged by loss-frequency assays [271]. As previously described, usage of the ParA-GFP and the *tetO*-TetR-mCherry labeling system also does not affect plasmid stability, indicating full functionality [156, 159]. ParB-GFP forms foci that are regularly positioned along the long cell axis *in vivo* (Figure 4.1A), consistent with ParB-binding to plasmid-encoded *parC* regions [170, 171].

Since plasmid dynamics occur primarily over the nucleoid, we reasoned that plasmid positioning with respect to the nucleoid rather than cell length is most informative. Therefore we measured ParB-GFP foci localization, together with Hoechst (DNA) stain to determine the nucleoid boundaries. Using the MATLAB-based software suite MicrobeTracker (MT) [272], we determined *E. coli* cell outlines from phase contrast (PC) images. Linear projections of ParB-GFP and Hoechst signal distributions along the long cell axis were also determined using the MT software package. The positions of the half-maxima of the linear Hoechst signal distribution in every cell were then determined. We defined the nucleoid length as the length between the two half-maxima of the Hoechst stain. The cell out-

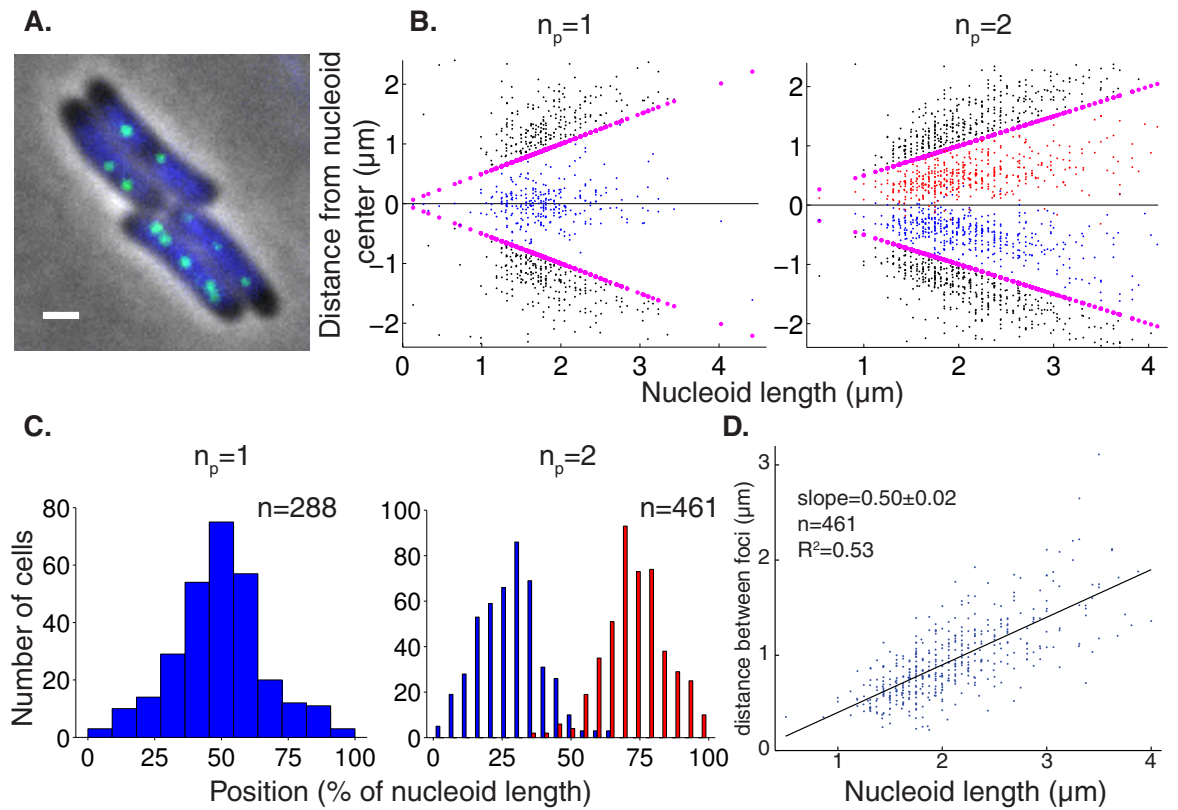


Figure 4.1: Plasmid foci are equally spaced over the nucleoid. (A) Fluorescence localization of plasmid-binding protein ParB-GFP (green) and Hoechst DNA stain (blue) in representative WT *E. coli* cells. Scale bars:  $1\mu\text{m}$ ; plasmid: pFS21 (mini-R1, *parC1+*, *parA+*, *parB::sfGFP*, *parC2+*). (B) Scatter plot of plasmid foci positions (blue, red) with respect to nucleoid edges (purple) and cell edges (black) for wild-type cells with  $n_p = 1, 2$  plasmid foci. (C) Histograms of plasmid foci positions shown in (B) relative to nucleoid length. (D) Scatter plot (blue) of the interplasmid focus distance as a function of nucleoid length in cells exhibiting two plasmid foci. A least square fit (black line) indicates a slope of 0.5. Experiments performed by Florian Szardenings.

lines were used together with the MATLAB tools spotFinderZ and spotFinderM [272] to determine ParB-GFP foci positions. The linear ParB-GFP distribution was used to control the peak detection method for false positives/negatives. This analysis allowed us to determine the positions of plasmid foci with respect to the nucleoid. As expected ParB-GFP foci colocalized exclusively with the Hoechst stain, and were equally spaced over the nucleoid (Figure 4.1B,C,D for  $n_p = 1, 2$ , Figure 4.2A,B for  $n_p = 3, 4$ ).

#### 4.2.2 *Mathematical analysis shows that dynamic, competitive ParA concentrations can generate equal plasmid spacing.*

Several studies have proposed that plasmid positioning is controlled by a concentration gradient of ParA over the nucleoid [166, 173, 176, 177, 180, 186]. Intuitively in this mechanism, ParB bound to plasmid *parC* (ParB-*parC* complex) interacts with nucleoid associated ParA-ATP, which effectively anchors the plasmid to the nucleoid. At the same time, the ParB-*parC* complex stimulates ParA-ATP hydrolysis causing a local ParA-ATP depletion. These processes could then generate a ParA-ATP gradient which a plasmid is able to follow. Reorganization of ParA gradients under the influence of multiple ParB-*parC* complexes might then lead to equal plasmid spacing. To rigorously understand if, and with what requirements, equal spacing can be achieved we develop here a minimal mathematical model based on the above principles.

We model the nucleoid as a 1d system of length  $L$  (along the long axis of the cell) on which ParA-ATP and plasmids can interact.  $A(x, t)$  denotes the nucleoid-associated ParA-ATP concentration at position  $x$  relative to one nucleoid edge at time  $t$ . Let  $x_1(t) \dots x_{n_p}(t)$  be the positions of the  $n_p$  plasmids. At each plasmid, ParA-ATP can be hydrolyzed with rate  $k_B$ , turning Par-ATP into a cytoplasmic ParA form, with copy number  $A_c(t)$ . After sufficiently long timescales, the cytoplasmic ParA becomes competent to bind the nucleoid again, with flux  $J(A_c(t))$ . Once bound to the nucleoid, ParA-ATP can diffuse along the nucleoid with diffusion constant  $D$ . This system can

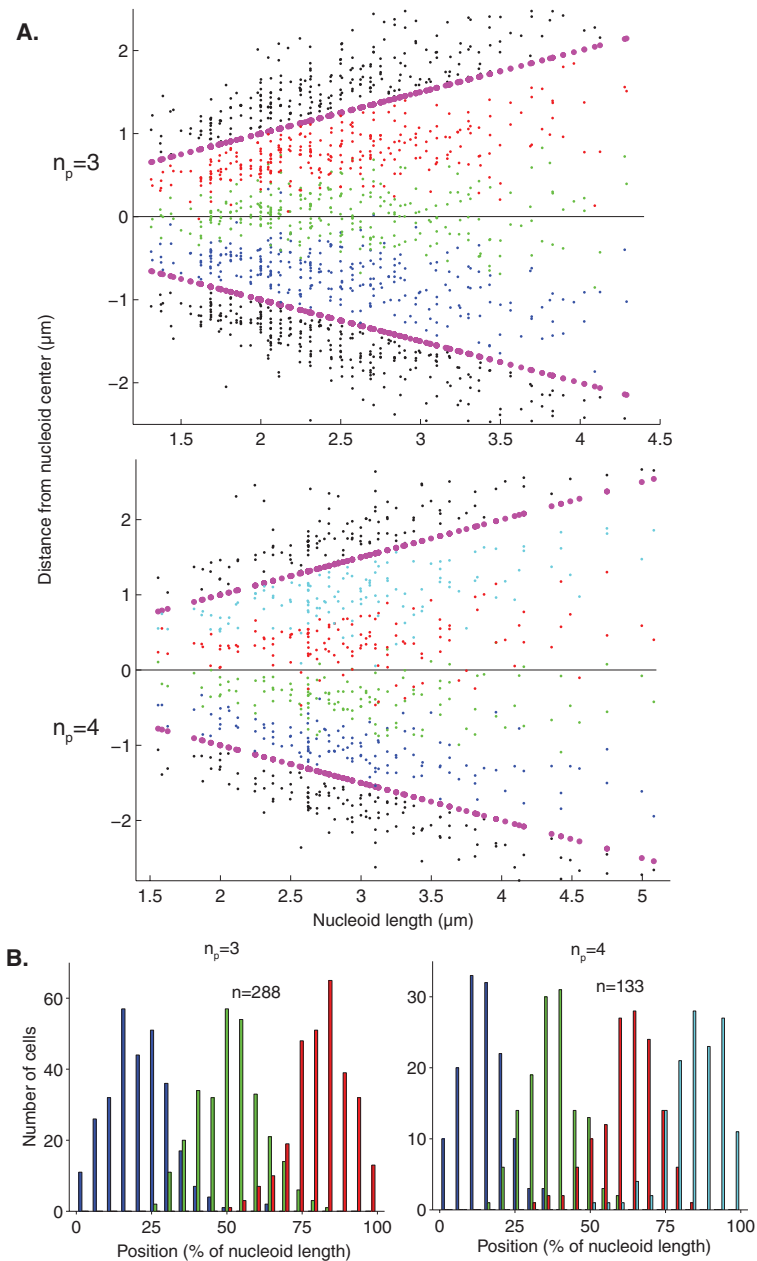


Figure 4.2: Plasmid foci are equally spaced over the nucleoid irrespective of nucleoid length or plasmid focus copy number. (A) Scatter plot of plasmid foci positions (blue, green, red, cyan) with respect to nucleoid edges (purple) and cell edges (black) for wild-type cells. Strains and plasmids used for Figure 4.2 are as described in Figure 4.1. (B) Histograms of plasmid foci positions shown in (A) relative to nucleoid length. Experiments performed by Florian Szardenings.

be described by the deterministic reaction-diffusion equations:

$$\begin{aligned} \frac{\partial A(x, t)}{\partial t} &= D \frac{\partial^2 A(x, t)}{\partial x^2} - k_B \sum_{i=1}^{n_p} A(x_i(t)) \delta(x - x_i(t)) + \frac{J(A_c(t))}{L} \\ \frac{dA_c(t)}{dt} &= k_B \sum_{i=1}^{n_p} A(x_i(t)) - J(A_c(t)) \\ \text{Boundary Conditions : } \frac{\partial A(x, t)}{\partial x} \Big|_{x=0} &= 0 = \frac{\partial A(x, t)}{\partial x} \Big|_{x=L} \text{ for all } t. \end{aligned} \quad (4.1)$$

Assuming a given (time-independent) total ParA copy number  $A_T$  in the system, we find that  $J$  is determined implicitly:

$A_T = A_c(t) + \int_0^L A(J(A_c(t)), x, t) dx$ . Note that the dimensions of  $k_B$  are length/time due to the dimensionality of the Dirac delta function  $\delta(x)$ .

For simplicity, we first assume that the ParA-ATP concentration at each plasmid is zero due to a high ParA-ATP hydrolysis rate  $k_B$ . Later on we will relax this assumption. We now use separation of time scales to obtain the steady-state solution for  $A(x)$ : we assume that plasmid motion is much slower than the time for individual ParA-ATP molecules to diffuse over the nucleoid and generate a concentration profile. In this way, the plasmid positions  $x_1 \dots x_{n_p}$  are time-independent and *a priori* unknown. Also  $J(A_c)$  is now time-independent and effectively a constant  $J$ . The equation for Equation 4.1 then simplifies to:

$$\begin{aligned} \frac{d^2 A(x)}{dx^2} &= -\frac{J}{LD} \\ \text{Boundary Conditions : } A(x_i) &= 0 \text{ for } 1 \leq i \leq n_p \\ \frac{dA(x)}{dx} \Big|_{x=0} &= 0 = \frac{dA(x)}{dx} \Big|_{x=L} \end{aligned} \quad (4.2)$$

Equation 4.2 can be solved for  $A(x)$  by integrating twice using the boundary conditions. The solution is given by:

$$A(x) = \begin{cases} \frac{J}{2LD} (x_1^2 - x^2) & \text{if } 0 \leq x \leq x_1, \\ \frac{J}{2LD} (-x^2 + (x_i + x_{i+1})x - x_i x_{i+1}) & \text{if } x_j \leq x \leq x_{j+1}, 1 \leq j < n_p, \\ \frac{J}{2LD} ((L - x_{n_p})^2 - (L - x)^2) & \text{if } x_{n_p} \leq x \leq L. \end{cases} \quad (4.3)$$

Next we use these equations to compute the diffusive fluxes of ParA-ATP,  $j_i = \left| D \frac{dA}{dx} \right|$ , at a plasmid location  $x_i$ , where the + and superscripts



below refer to the flux from the right (+) and left (−) respectively. We find:

$$\begin{aligned} j_1^- &= \frac{Jx_1}{L}, \\ j_i^+ &= j_{i+1}^- = \frac{J(x_{i+1} - x_i)}{2L}, \\ j_{n_p}^+ &= \frac{J(L - x_{n_p})}{L}. \end{aligned}$$

Clearly, a symmetric ParA concentration profile, where fluxes from either side balance, is only possible for  $x_1 = L - x_{n_p} = \frac{1}{2}(x_{i+1} - x_i)$ . The plasmids are then equally distributed with

$$x_j = \frac{L}{2n_p} + \frac{L}{n_p}(j - 1).$$

We note that the predicted inter-plasmid spacing  $\frac{L}{n_p}$  arising from this analysis is consistent with our experimental findings (Figure 4.1D, Figure 4.2B).

Importantly, the above analysis provides insight into the equal spacing mechanism. The key is that the above fluxes depend on the distances either between the plasmid and nucleoid end, or between neighboring plasmids. This feature is a consequence of ParA binding to the nucleoid anywhere, but with ParA release only occurring at a plasmid. In order for these on and off fluxes to balance at steady-state, the off-flux at a plasmid must scale with the inter-plasmid or plasmid-nucleoid-end distance. In this way, non-local information about lengths is converted into local spacing information encoded in the slope of ParA-ATP concentration. For non-equal plasmid spacing, the competing ParA concentrations on either side of a plasmid will be unequal, with one gradient steeper than the other. The steeper gradient corresponds to the side with the greater available space for ParA binding. If a plasmid can preferentially move (on the appropriate slow time scale) towards the side with the locally steepest ParA-ATP concentration, the plasmids are then progressively restored towards equal spacing. As this process occurs, the ParA-ATP concentrations will dynamically reorganize such that a symmetric configuration around a plasmid is reached only when the plasmids are equally spaced. In this state, where the competing ParA-ATP concentrations are symmetric, plasmid movement would no longer

have a directional preference and would thus remain, on average, stationary.

So far, we have assumed that the ParA-ATP concentration vanishes at a plasmid, corresponding to very fast ParA-ATP hydrolysis. However, our results also hold true when we only assume that this hydrolysis occurs with a finite rate  $k_B$ , leading to a non-zero concentration of ParA-ATP at a plasmid. This ParA-ATP can then anchor a plasmid to the nucleoid before being hydrolysed. This more general and realistic case is presented in the Materials and Methods section, but our overall conclusions reached above remain unchanged.

From the above analysis, we see that the following conditions are required for equal plasmid spacing: (1) movement of a plasmid towards higher ParA-ATP concentrations. (2) diffusion of (at least a fraction of) ParA-ATP over the nucleoid to ensure formation of competitive concentration gradients. Single molecule tracking experiments in vitro support this assumption [177, 180]. (3) ParA-ATP hydrolysis must occur (predominantly) by plasmid-associated ParB-*parC* complexes, again to ensure gradient formation. (4) ParA-ATP must adopt a 1d-like configuration, as previously claimed [157–159]. If ParA were not organized in this fashion, it would be possible for ParA to diffuse around the sides of a plasmid without encountering the hydrolyzing effect of the ParB-*parC* complex. This would equalize the ParA concentrations on both sides even in the case of asymmetrically placed plasmids, leading to failure of the equal spacing mechanism. This assumption is in line with our subsequent experiments (see below). Due to this proposed 1d-like nature, we will from now on refer to the ParA distributions away from a plasmid as ParA structures. (5) There must be a separation of time scales between plasmid movement and ParA concentration reorganization, as discussed above.

Importantly, this overall mechanism is not reliant on a specific type of plasmid translocation. Any process that would allow a plasmid to move into regions of higher ParA concentration will suffice. In the following sections we therefore analyze different means of plasmid

movement and compare them with our experimental data to determine which is used in our *par2* segregation system.

#### 4.2.3 *Diffusion/immobilization model could space highly mobile plasmids equally over the nucleoid.*

In the previous section the mechanistic details of plasmid movement towards a higher ParA concentration were not specified. We now examine a specific implementation involving a diffusion-immobilization mechanism. Using a minimal modelling approach, we assume that nucleoid-associated ParA-ATP can immobilize freely diffusing plasmids through its interaction with the ParB-*parC* complex and that ParA-ATP does not polymerize (Figure 4.3A). Since the plasmid will tend to become immobilized in regions of higher ParA-ATP concentration, this process allows for effective plasmid translocation up a ParA-ATP concentration gradient. We also incorporate ParB-*parC*-stimulated ParA-ATP hydrolysis at a plasmid, in accordance with prior experimental data. Here we use standard diffusion for the plasmid movement; below we discuss the potential impact of subdiffusive motion. To further investigate this mechanism, given the known physiological and biochemical constraints, we developed stochastic simulations, implemented in C++, using a Gillespie algorithm [218].

In our simulation, a one dimensional lattice with sites of size  $dx = 5nm$  represents the nucleoid, sites are numbered  $0...(L - 1)$ . The reactions and corresponding propensities  $p_t$  are fully described in Table 4.1. Parameter values used are listed in Table 4.2. Reactants are:

- $A_i$ : ParA-ATP at site  $i$  with number  $A[i] \geq 0$ ;
- $P_{j,i}$ : plasmids with  $j$  ParA-ATP bound to it at site  $i$  with number  $P[j][i](\geq 0)$ ;
- $A_{ADP}$ : cytoplasmic ParA-ADP with number  $AADP (\geq 0)$ ;
- $A_{cyto}$ : cytoplasmic ParA-ATP with number  $ACYTO (\geq 0)$ .

ParA-ATP and plasmids can diffuse on the lattice with diffusion coefficient  $D_A$  and  $D_P$  respectively. Up to 35 ParA-ATP can bind to a plasmid at the same site with reaction parameter  $k_{AB}$  reflecting the binding interaction of ParA-ATP and the ParB-*parC* complex [171].

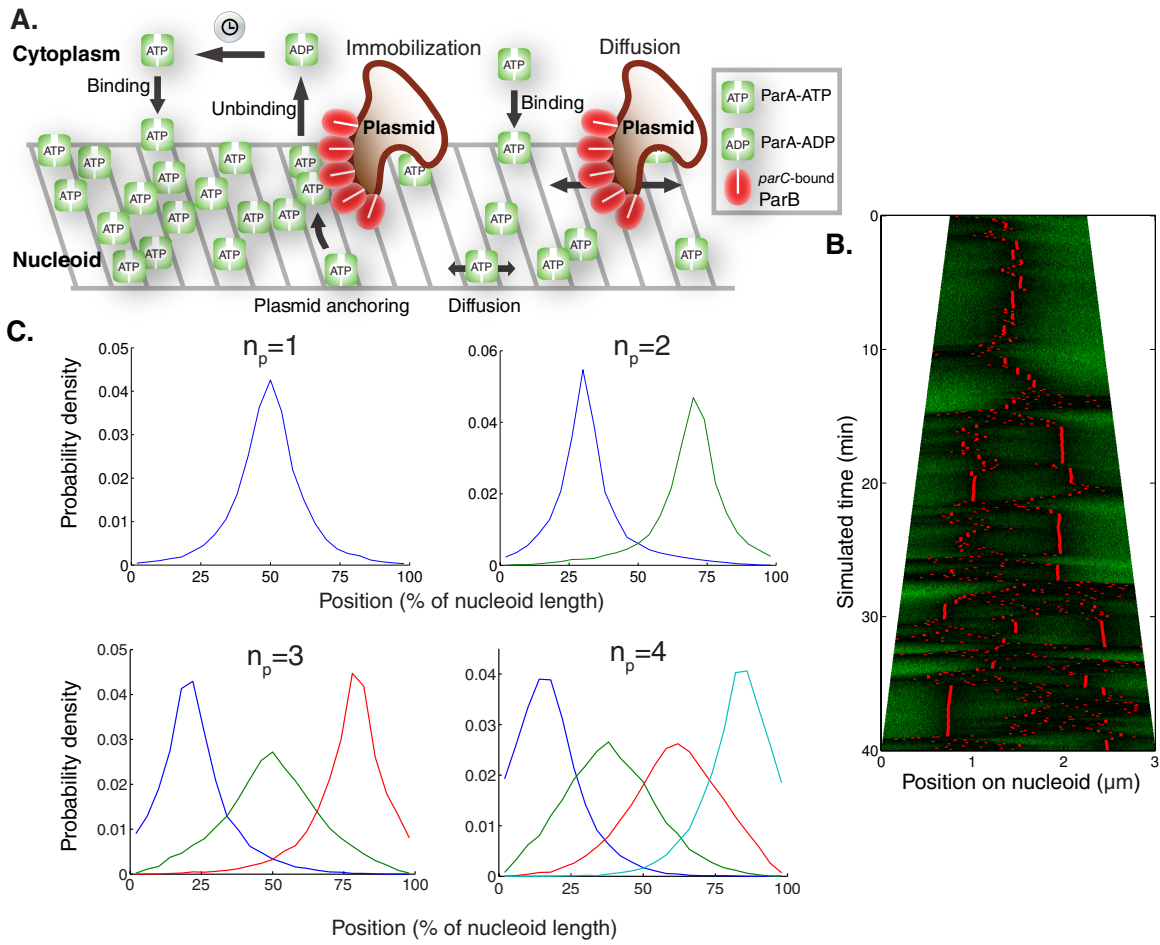


Figure 4.3: Diffusion/immobilization model can move and maintain plasmids at equally spaced positions. (A) Schematic illustration of *par2* diffusion/immobilization model. The clock indicates the slow conversion of cytoplasmic ParA-ADP into cytoplasmic ParA-ATP that is competent to bind to the nucleoid. The corresponding reactions and propensities  $p_t$  are fully described in Table 4.1. (B) Typical simulation kymograph of diffusion/immobilization model for growing cell, where plasmid (red) diffusion influenced by the local ParA-ATP (green) concentration leads to immobilization initially at mid cell. After plasmid duplication, the system dynamically self-organizes to reacquire equal plasmid spacing. (C) Time-averaged plasmid position distributions for diffusion/immobilization model with  $n_p = 1 - 4$  on a simulated nucleoid growing from  $1.5\mu\text{m}$  to  $3\mu\text{m}$  in  $40\text{min}$  without plasmid duplication. Plasmid distributions were obtained by sampling positions every 5s in 36 independent simulations.

Description	Reactions	Propensities $p_t$
ParA-ATP diffusion	$A_i \rightarrow A_{i+1}, i = 0 \dots (L-2)$	$\frac{D_A}{dx^2} \cdot A[i]$
	$A_{i+1} \rightarrow A_i, i = 0 \dots (L-2)$	$\frac{D_A}{dx^2} \cdot A[i+1]$
Plasmid diffusion / immobilization	$P_{j,i} \rightarrow P_{j,i+1}, i = 0 \dots (L-2), j = 0 \dots 35$	$\frac{D_P}{dx^2} \cdot P[j][i]$ if $j = 0$ $\frac{D_A}{dx^2} \cdot P[j][i]$ if $j = 1$ 0 if $j > 1$
	$P_{j,i+1} \rightarrow P_{j,i}, i = 0 \dots (L-2), j = 0 \dots 35$	$\frac{D_P}{dx^2} \cdot P[j][i+1]$ if $j = 0$ $\frac{D_A}{dx^2} \cdot P[j][i+1]$ if $j = 1$ 0 if $j > 1$
Plasmid anchoring	$P_{j,i} + A_i \rightarrow P_{j+1,i}, i = 0 \dots (L-1), j = 0 \dots 34$	$k_{AB} \cdot P[j][i] \cdot A[i]$
ParA-ATP unbinding from plasmid	$P_{j+1,i} \rightarrow P_{j,i} + A_{ADP},$ $i = 0 \dots (L-1), j = 0 \dots 34$	$k_B \cdot P[j+1][i]$
Cytoplasmic ParA-ADP to ParA-ATP conversion	$A_{ADP} \rightarrow A_{cyto}$	$k_W \cdot A_{ADP}$
ParA-ATP nucleoid binding	$A_{cyto} \rightarrow A_i, i = 0 \dots (L-1)$	$\frac{k_{on}}{L} \cdot A_{CYTO}$

Table 4.1: Reactions and propensities used in the diffusion/immobilization model

More than one ParA-ATP bound to a plasmid reduces the plasmid diffusion constant to zero. We varied the exact number of ParA-ATP molecules forming a complex that are required to completely immobilize the plasmid and this variation does not alter the qualitative behavior of the system. Plasmid-bound ParA-ATP can be hydrolysed with reaction parameter  $k_B$ . Whenever a ParA-ATP hydrolysis event occurs, ParA unbinds from the nucleoid and becomes a cytoplasmic ParA-ADP. ParA-ADP can then be converted into a cytoplasmic ParA-ATP that is competent in DNA binding (cytoplasmic ParA-ATP for short) with a slow reaction parameter  $k_W$  [166]. We do not keep track of the spatial positions of ParA-ADP and ParA-ATP in the cytoplasm. Instead we merely keep track of their number. Introduction of a low spontaneous ParA-ATP hydrolysis parameter  $k_{off}$  also does not alter the behaviour of the system. Cytoplasmic ParA-ATP can then bind anywhere along the nucleoid with parameter  $k_{on}$ . See subsection 4.4.4 for further details on the simulation duration and model outputs.

Prior work has demonstrated plasmid displacement along the long

Parameter	Description	Value	Notes
$D_A$	Nucleoid bound ParA-ATP diffusion constant	$10^{-2}\mu\text{m}^2/\text{s}$	Fitted, can be increased without loss of qualitative behaviour of system.  Nevertheless, it is difficult to physically reconcile more mobile nucleoid-bound ParA-ATP with the ability to immobilize a plasmid with a lower diffusion constant.  Therefore we have assumed that ParA-ATP diffuses 10x slower than the plasmid, ensuring that the assumption that ParA-ATP can immobilize plasmids is physically justified.
$D_P$	Plasmid diffusion constant	$10^{-1}\mu\text{m}^2/\text{s}$	A relatively high value is needed for compatibility with previous experiments[159]. An upper bound on the plasmid diffusion constant from experiments (Figure 4.4C) turned out to be too low for this model to fit our experimental observations.
$k_{on}$	ParA-ATP nucleoid binding	$50\text{s}^{-1}$	Constrained by experiment [166].
$k_{AB}$	ParA-ATP to plasmid binding	$100\text{s}^{-1}$	Fitted, should be high enough to allow for plasmid immobilization.
$k_B$	Plasmid bound ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$68.5\text{s}^{-1}$	Fitted together with $D_A$ and $k_W$ to ensure equal plasmid spacing.
$k_W$	(Cytoplasmic) ParA-ADP to ParA-ATP conversion	$1/15\text{s}^{-1}$	Constrained by experiment [166],  this value should be low enough to ensure that  cytoplasmic ParA diffusion can generate a uniform cytoplasmic  ParA-ATP and ParA-ADP concentration.

Table 4.2: Parameter values used in the diffusion/immobilization model

cell axis of up to  $3 - 4\mu\text{m}$  within  $10\text{min}$  [159, 183]. With a diffusion/immobilization mechanism all plasmid movement in between immobilization events is generated by (unbiased) free diffusion, for which we have (in 1d) a mean square displacement (MSD) of  $\langle r^2(t) \rangle = 2D_P t$ . By inserting the above length and time scales into this equation, we conclude that a plasmid diffusivity of at least  $D_P \sim 10^{-2}\mu\text{m}^2\text{s}^{-1}$  would be required to generate sufficiently rapid diffusive movement in accordance with previous experiments. We therefore chose  $D_P = 10^{-1}\mu\text{m}^2\text{s}^{-1}$ . In order to physically justify that ParA can immobilise the plasmids, we chose the nucleoid bound ParA-ATP diffusivity to be lower than  $D_P$ , with  $D_A = 10^{-2}\mu\text{m}^2\text{s}^{-1}$  (Table 4.2). We experimentally constrained the overall copy number of ParA for pB171 *par2* by semi-quantitative Western blots, which revealed that there were approximately  $8 \cdot 10^3$  ParA monomers per cell [271]. This diffusion/immobilization model could produce equal plasmid spacing

on simulated growing nucleoids with varying numbers of plasmids (Figure 4.3B,C). This result demonstrates that using a sufficiently high (low) plasmid (ParA) diffusivity, respectively, the equal plasmid spacing seen in our experiments (Figure 4.1B,C,D, Figure 4.2A,B) and previously [159], could in principle be achieved using a diffusion/immobilization mechanism.

#### 4.2.4 *Free plasmid mobility is too low for a diffusion/immobilization mechanism.*

To test whether the requirement of a relatively high free plasmid mobility is met *in vivo*, we compared the movement of test-plasmids with and without *par2*. We analyzed trajectories of labeled plasmid foci using the *tetO*-TetR-mCherry labeling system, measuring the positions over time (Figure 4.4A) and MSDs for each time lag  $\tau$ . Plasmid motion will be biased by a functional *par2*<sup>+</sup> partitioning system, in contrast to the random motion of *par*<sup>-</sup>. Nevertheless comparing MSDs can still be informative in comparing relative overall mobilities.

Cell outlines and linear projections of *tetO*-TetR-mCherry and Hoechst signal distributions along their long axis were determined as described in subsection 4.2.1, as well as the distribution of *tetO*-TetR-mCherry-labeled plasmids along the long axis of cells. *tetO*-TetR-mCherry foci detection was also performed using the methods described above in *par*<sup>-</sup> time-lapses of 1min (short) or 15min (long) in duration with images taken at intervals of 4s or 30s respectively. For the short time-lapses we analysed cells with one or more foci, although all our results were unchanged if analysis was restricted to one focus cells to prevent potential foci labelling errors. For the long time-lapses, we only analysed cells exhibiting one focus. This was due to difficulties in distinguishing between multiple foci due to merging/splitting events, out of focus plane movement and photobleaching when acquiring images using a time interval of 30s. These effects could have resulted in biases in the analysis due to labelling errors. We were unable to lower the time interval and simultaneously image for long time periods due to TetR-mCherry photobleaching.

At every time point the two-dimensional squared foci displacements  $r^2(\tau)$  after time lag  $\tau$  were determined. All measured displacements for the same time lag were then averaged together to obtain the MSD  $\langle r^2(\tau) \rangle$  with time lags from 4s to 15min (Figure 4.4B,C, Figure 4.5A).

On time scales up to a minute we found that the  $par2^+$  MSD is higher than in  $par^-$  (Figure 4.4B), showing that, on average,  $par2^+$  plasmids are more mobile than their  $par^-$  counterparts. Note that the number of data points for the short time lags far exceeds the number of trajectories ( $n_{par^-} = 747$ ,  $n_{par2^+} = 763$ ), since every trajectory contains multiple short time lags. Consequently our estimates for the mean are relatively precise for short time lags. It is true that the error on the mean does not reflect inaccuracy due to experimental limitations in determining the actual plasmid position. The measured plasmid displacement  $r(\tau)$  can report the true plasmid displacement  $r_p(\tau)$  at a resolution no greater than our measurement error, which can be up to  $0.1\mu m$  due to microscope drift. Our measurements are also limited by a finite pixel size of  $0.066\mu m$ . We therefore have:  $r(\tau) = r_p(\tau) + \varepsilon$ , where  $\varepsilon$  is the error due to both of the above effects. Squaring and averaging over many plasmid trajectories results in an MSD:  $\langle r^2(\tau) \rangle = \langle r_p^2(\tau) \rangle + \langle \varepsilon^2 \rangle + \langle 2\varepsilon r_p(\tau) \rangle$ . The last term vanishes due to averaging, but the second term remains and generates a small time independent value for  $\tau > 0$ . However, that error is the same for both  $par2^+$  and  $par^-$ .

Overall, these results are hard to reconcile with a diffusion/immobilization mechanism where the  $par2$  system can only immobilize plasmids, and thus lower their MSD. These MSD values could in principle be limited due to cellular confinement. However, we found that MSD saturation only starts to occur at much larger length scales at times of up to 10min (Figure 4.4C).

In the presence of  $par2$ , plasmids generally reside within the nucleoid region, while in its absence they tend to become somewhat more polar localized, although they can still sample the entire cell volume on long enough timescales [199]. Consistently we still find many  $par^-$  plasmids located within the nucleoid region (Figure 4.5A). Restrict-



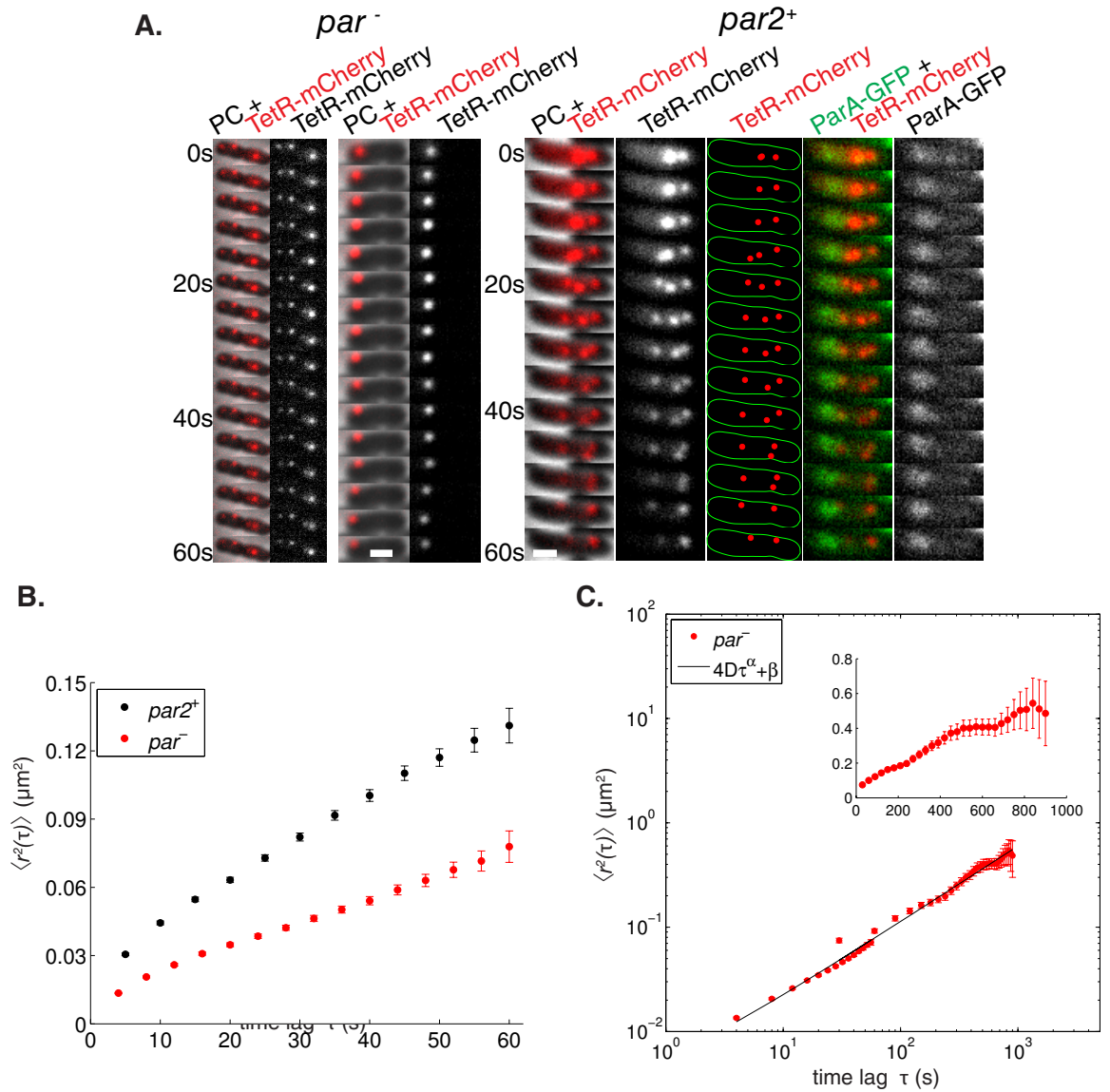


Figure 4.4: The *par2* segregation system increases plasmid mobility. (A) Time lapses showing the localization of *par*<sup>-</sup> pMH82tetO120 (mini-R1, *par*<sup>-</sup>, *tetO120*) and *par*<sup>2+</sup> pSR236 (mini-R1, *parC1*<sup>+</sup>, *parA*<sup>-</sup>, *parB*<sup>+</sup>, *parC2*<sup>+</sup>, *tetO120*, *P<sub>lac</sub>::parA::eGFP*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). The *par*<sup>2+</sup> time lapse, with ParA-GFP localization, shows a segregation event where two foci segregate  $\geq 0.8\mu\text{m}$  further apart within 20s. PC=phase contrast, scale bar:  $1\mu\text{m}$ . (B) Mean square displacements after time lag  $\tau$  were extracted from plasmid trajectories ( $n_{\text{par-}} = 747$ ,  $n_{\text{par2+}} = 763$ ) using strains specified in (A), *par*<sup>-</sup> (red) and *par*<sup>2+</sup> (black), error bars: standard error of the mean. (C) Log-log plot of experimental mean square displacements after time lag  $\tau$  (red) were extracted from plasmid trajectories over 1 min as in (A,B) and (inset, linear scales,  $n = 50$ ) over 15min from *par*<sup>-</sup> pMH82tetO120 (mini-R1, *par*<sup>-</sup>, *tetO120*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). At timescales on the order of 10min saturation of the MSD occurs due to cellular confinement. A nonlinear least square fit (black line) using the function was used to estimate parameter values:  $\alpha = 0.73 \pm 0.02$ ,  $D_p = 9.7 \pm 1.3 \cdot 10^{-4} \mu\text{m}^2 \text{s}^{-\alpha}$ ,  $\beta = 1.6 \pm 2.4 \cdot 10^{-3} \mu\text{m}^2$ , ( $R^2 = 0.99$ , p-values:  $8 \cdot 10^{-15}$ ,  $8 \cdot 10^{-3}$  and 0.50 respectively). Error bars: standard error of the mean. Experiments performed by Florian Szardenings.

ing the mobility analysis to  $par^-$  plasmids within the nucleoid region did not alter the resulting MSD curves significantly (Figure 4.5A). We conclude that the presence of  $par2$  can increase plasmid mobility in the nucleoid region, which is inconsistent with a diffusion/immobilization mechanism. We emphasize that this conclusion can be made irrespective of the underlying ( $par^-$ ) plasmid transport processes, which we now describe in more detail.

It has been reported that chromosomal loci and RNA-protein particles exhibit subdiffusive, rather than diffusive, behavior in the cytoplasm [273, 274]. Therefore it is possible that plasmids without a segregation mechanism could also exhibit subdiffusive motion. Further analysis is required to fully distinguish subdiffusion from the additional effects of cellular confinement or glass-like properties of the bacterial cytoplasm [199, 274]. Nevertheless such additional analysis is not required for the conclusions on  $par^-$  plasmid mobility relevant to this study, as we now explain. Subdiffusion results in an expected MSD displacement of the form  $\langle r^2(t) \rangle = 4D_p t^\alpha$ , with  $\alpha < 1$  and  $D_p$  the apparent diffusion constant (in units of  $\mu m^2 s^{-\alpha}$ ). Even at short timescales of up to a minute, the MSD has a nonlinear shape, as has been reported before [161]. This is fully consistent with subdiffusive motion on these timescales. Taking into account our time lag independent measurement error, we thus expect the experimentally observed planar MSD for free particle subdiffusion in three dimensions to have the form:

$$\langle r^2(t) \rangle = 4D_p t^\alpha + \beta.$$

We performed a nonlinear least squares fit (weighted by the standard error of the mean (SEM):  $1/SEM(\tau)$ ) resulting in the values  $\alpha = 0.78 \pm 0.04$ ,  $D_p = 6.8 \pm 1.2 \cdot 10^{-4} \mu m^2 s^{-\alpha}$ ,  $\beta = 6 \pm 1 \cdot 10^{-3} \mu m^2$  ( $R^2 = 0.99$ , p-values:  $4 \cdot 10^{-10}$ ,  $1 \cdot 10^{-4}$  and  $8 \cdot 10^{-4}$  respectively). On longer timescales up to 15min (Figure 4.4C), plasmid mobility also showed subdiffusive behaviour with a similar analysis giving  $\alpha = 0.78 \pm 0.05$ ,  $D_p = 6.2 \pm 2.1 \cdot 10^{-4} \mu m^2 s^{-\alpha}$ ,  $\beta = 4 \pm 1 \cdot 10^{-2} \mu m^2$  ( $R^2 = 0.99$ , p-values:  $8 \cdot 10^{-15}$ ,  $8 \cdot 10^{-3}$  and  $2 \cdot 10^{-3}$  respectively). Analysing the two datasets combined (Figure 4.4C) also generated consistent results, although the constant  $\beta$  was not significantly different from zero in this case:  $\alpha = 0.73 \pm 0.02$ ,  $D_p = 9.7 \pm 1.3 \cdot 10^{-4} \mu m^2 s^{-\alpha}$ ,  $\beta = 1.6 \pm 2.4 \cdot$

$10^{-3} \mu m^2$ , ( $R^2 = 0.99$ , p-values:  $8 \cdot 10^{-15}$ ,  $8 \cdot 10^{-3}$  and 0.50 respectively, fit shown in Figure 4.4C). Fitting  $\langle r^2(t) \rangle = 4D_p t^\alpha$  instead to this combined data set did not alter our estimates for  $\alpha$  and  $D_p$  significantly. We conclude that our MSD displacements on both short and long timescales are well described by subdiffusion with  $\alpha = 0.7 - 0.8$  and an apparent diffusion constant  $D_p = 5 - 10 \cdot 10^{-4} \mu m^2 s^{-\alpha}$  (Figure 4.4C). This is consistent with other recent reports on *par*<sup>-</sup> plasmid mobility [161, 199].

On all observable timescales (i.e. 4s and longer) the experimentally found *par*<sup>-</sup> MSD is bounded from above by the function  $4D_f \tau$ , with  $D_f = 10 \cdot 10^{-4} \mu m^2 s^{-1}$ . Moreover, free diffusion with diffusion constant  $D_f$  inside a box of cellular dimensions still exceeds the experimental subdiffusive mobility (data not shown). So the experimental MSD is lower on all observed timescales than a hypothetical particle that would perform free diffusion inside a cell with a diffusion constant  $D_f$ . This upper limit is already much lower than that needed to be consistent with the previously reported plasmid displacement data discussed above. We will further exploit this upper limit in our analysis below.

To further investigate the effect of *par2* on plasmid positioning, we also studied rapid plasmid segregation events. We defined these as cases where two plasmid foci whose separation is initially  $\leq 0.3 \mu m$ , move within 20s at least another  $0.8 \mu m$  apart (Figure 4.4A, Figure 4.5B). We also allowed for the two foci to be initially merged. Using these criteria, despite equally large data sets, we found 13 such events in *par2*<sup>+</sup> and only one such case in *par*<sup>-</sup>. Furthermore, we only retrieved 2 further *par2*<sup>+</sup> segregation events when we relaxed the criterion to separation within 60 s instead of 20 s. This analysis shows that most segregation events occur rapidly. When we investigated the 26 plasmid trajectories involved they showed larger maximal MSDs compared to sets of 26 trajectories that were repeatedly randomly sampled from the whole *par2*<sup>+</sup> dataset ( $p < 10^{-6}$ ). This finding indicates that the *par2* system can particularly enhance the mobility of plasmids when they are in close proximity. We then simulated 300 plasmid duplication events with our diffusion/immobilization

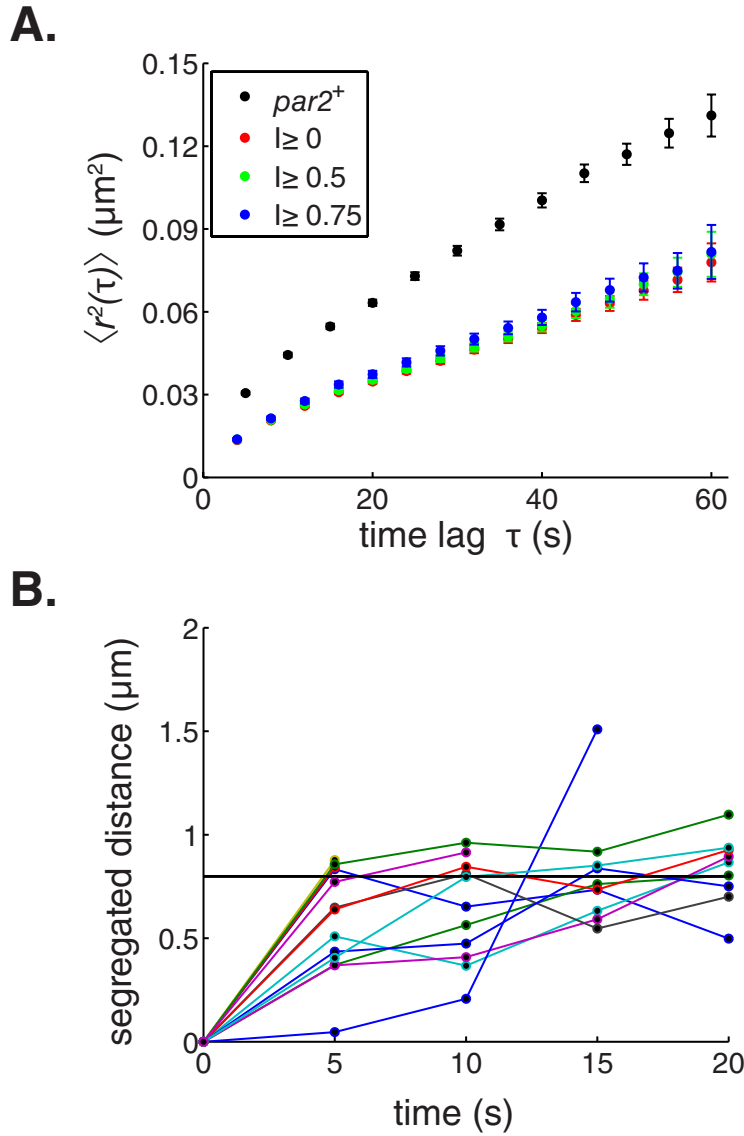


Figure 4.5: Plasmid mobility analysis in absence and presence of *par2* segregation system. (A) Plots as in Figure 4.4B except with experimental *par*<sup>−</sup> (red, green, blue) plasmid trajectories in which plasmid location is within a region of normalized Hoechst stain intensity *I* equal to or higher than the values indicated in the legend. The corresponding plasmid copy numbers ( $n_{par2+} = 763$ ,  $n_{par-, I \geq 0} = 747$ ,  $n_{par-, I \geq 0.5} = 592$ ,  $n_{par-, I \geq 0.75} = 401$ ) indicate that a large fraction of *par*<sup>−</sup> plasmids do indeed reside in the nucleoid region; error bars: standard error of the mean. (B) Plots of 13 segregation events of *par*<sup>+</sup> pSR236 (mini-R1, *parC1*<sup>+</sup>, *parA*<sup>−</sup>, *parB*<sup>+</sup>, *parC2*<sup>+</sup>, *tetO120*, *P<sub>lac</sub>::parA::eGFP*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). Shown is the additionally segregated distance (colored lines) as a function of time, both with respect to the start of each segregation event. A segregation event is defined as two foci that are initially  $\leq 0.3 \mu\text{m}$  apart and subsequently segregate  $\geq 0.8 \mu\text{m}$  further apart within 20s. The horizontal line (black) indicates  $0.8 \mu\text{m}$ . Experiments performed by Florian Szardenings.

model to determine the magnitude of diffusion constant required to generate the experimentally observed segregation. Note that we used diffusion rather than subdiffusion here because we have already determined that *par*<sup>-</sup> plasmid movement is slower on all observed timescales than free diffusion with a diffusion constant  $D_f = 10 \cdot 10^{-4} \mu m^2 s^{-1}$ . Hence, if the required diffusion constant is larger than  $D_f$  then we have also ruled out a subdiffusion/immobilization model. We required that 5% (15 out of 300) of segregated distances within 20 s were at least  $0.8 \mu m$  (a very conservative requirement, since the criterion was satisfied by 13 of our 15 experimental segregation events). This requirement necessitated a free plasmid diffusion constant on the order of  $10^{-1} \mu m^2 s^{-1}$ , about two orders of magnitude higher than our experimentally observed upper bound  $D_f$  on the experimental *par*<sup>-</sup> plasmid mobility. Hence, we conclude that the plasmids are generally too immobile for a diffusion/immobilization (or subdiffusion/immobilization) mechanism to explain these segregation events. Also the qualitative behaviour of segregation events in the diffusion/immobilization model appears different, since experimental segregation events (Figure 4.4A, Figure 4.5B) show more directionally biased motion, while the diffusion/immobilization model generates more sustained random, diffusive motion during segregation, prior to immobilization at equally spaced positions (Figure 4.3B). Nevertheless, these segregation events were sufficiently rare not to significantly alter the overall MSD behaviour of the entire dataset shown in Figure 4.4B. Thus the increased average mobility in the presence of *par2*<sup>+</sup> cannot only be ascribed to these segregation events.

It is possible that the *tetO*-TetR-mCherry labeling system caused reduced plasmid mobility as compared to unlabelled plasmids. However, as we used the same labeling method for both *par2*<sup>+</sup> and *par*<sup>-</sup> cases, our above conclusions on relative mobility are unaffected. Moreover, our *tetO*-TetR-mCherry labeled plasmids still exhibited rapid segregation events (such as in Figure 4.4A), underscoring the ability of *par2* to overcome low plasmid mobility. Overall, we find that diffusion/immobilization cannot explain our data on *par2*<sup>+</sup> versus *par*<sup>-</sup> plasmid mobility, as well as on rapid *par2*<sup>+</sup> plasmid segregation.

#### 4.2.5 *ParA structures competing to direct plasmid motion can space plasmids equally over the nucleoid.*

Given the shortcomings of the diffusion/immobilization model, we next tested models based on directed motion, allowing more rapid directed rather than unbiased diffusive plasmid movement. More specifically, we tested models based on the formation of competing ParA polymers, with ParB-*parC* -stimulated ParA-ATP hydrolysis directing plasmid movement. By modulating the length of these polymers, we thereby tested the robustness of directed motion models to generate equal plasmid positioning.

We again used a Gillespie algorithm to simulate ParA dynamics on the nucleoid (see Figure 4.6A for illustration). The reactions and corresponding propensities  $p_t$  are described in Table 4.3. The nucleoid was represented as a rectangular lattice ( $dx = 5nm$  in both dimensions), with a much shorter width ( $30nm$ ) than length (several  $\mu m$ ). Thus every site had a coordinate along the long axis (labelled as  $0 \dots L-1$ ) as well as a coordinate along the short axis (labelled  $0 \dots S-1$ ).

Reactants are:

$Am_{i,j}$ : mobile ParA-ATP at site  $(i, j)$  with number  $A_m[i][j] \geq 0$  ;

$A_{i,j}$ : polymeric ParA-ATP at site  $(i, j)$  with number  $A[i][j] (= 0 \text{ or } 1)$ ;

$P_i$ : plasmids at site  $i$  with number  $P[i] (\geq 0)$ ;

$A_{ADP}$ : cytoplasmic ParA-ADP with number  $A_{ADP} (\geq 0)$ ;

$A_{cyto}$ : cytoplasmic ParA-ATP with number  $A_{CYTO} (\geq 0)$ .

Parameter values used are listed in Table 4.4.

Similar reactions as in the diffusion/immobilization model described the cytoplasmic dynamics of ParA-ADP and ParA-ATP (compare Table 4.1 and Table 4.3). Nucleoid-associated ParA-ATP could also still diffuse across the nucleoid in a mobile state in all four directions to neighbouring sites with diffusion constant  $D_A$ . However, two of these molecules at sites neighboring each other along the long nucleoid axis could interact to form a ParA polymer of two subunits, with reaction parameter  $k_p$ . Further ParA-ATP polymerization could occur by attachment of mobile ParA-ATP, located at a site immediately next to the tip of an existing ParA polymer, but only along

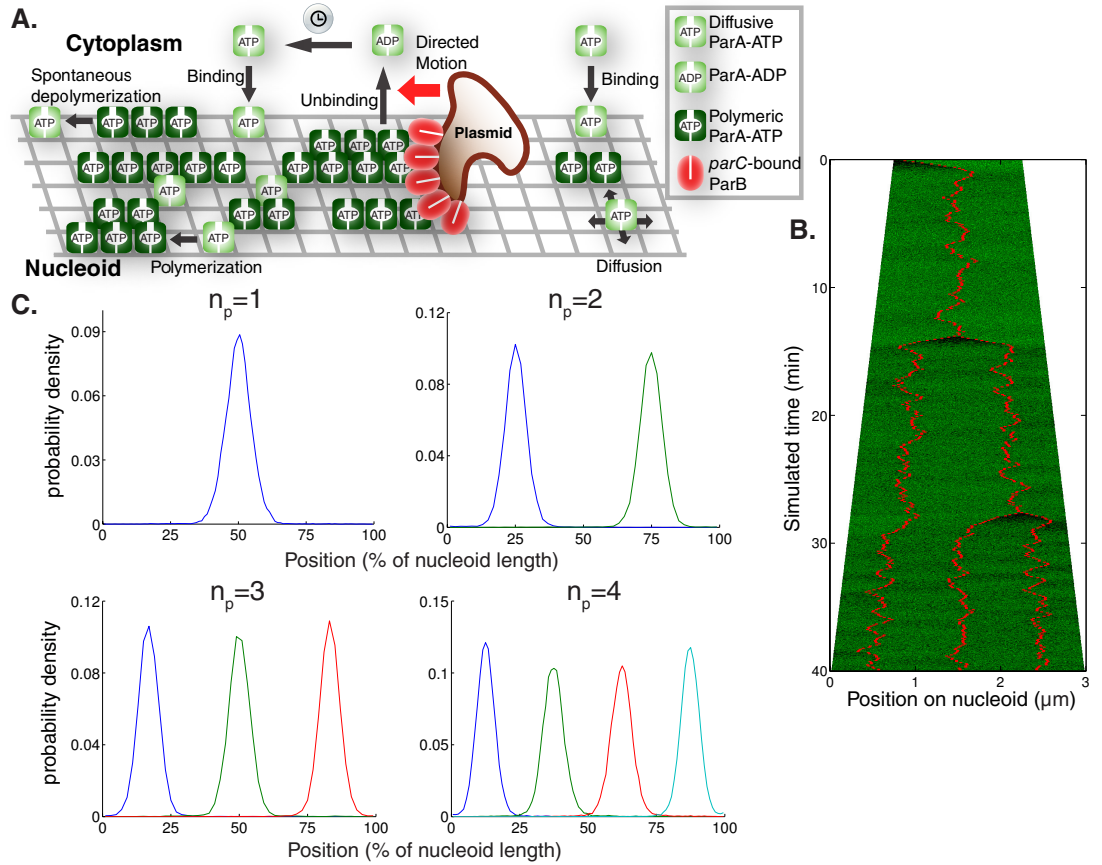


Figure 4.6: The directed motion model can move and maintain plasmids at equally spaced positions. (A) Schematic illustration of *par2* directed motion model. The clock indicates the slow conversion of cytoplasmic ParA-ADP into cytoplasmic ParA-ATP that is competent to bind to the nucleoid. The corresponding reactions and corresponding propensities  $p_i$  are described in Table 4.3. (B) Typical simulation kymograph of directed motion model with short polymers for a simulated growing cell where a plasmid (red) is initially directed from a nucleoid edge to mid-cell by ParA (green) filament competition. After plasmid duplication, the system dynamically self-organizes to attain equal spacing. (C) Time-averaged plasmid position distributions for directed motion model with short polymers for  $n_p = 1 - 4$  plasmids on a simulated nucleoid growing from  $1.5\mu\text{m}$  to  $3\mu\text{m}$  in  $40\text{min}$  without plasmid duplication. Plasmid distributions were obtained by sampling positions every  $5\text{s}$  in 36 independent simulations.



the long axis. ParA-ATP polymers were assumed to be immobile. A ParA-ATP polymeric subunit could depolymerize spontaneously with reaction parameter  $k_{dp}$ , i.e. be converted into a mobile ParA-ATP at the same site. Given that its size is similar to the width of the lattice, we only took into account the plasmid position along the long axis and we assumed that it occupied all sites along the short axis simultaneously. The plasmid could diffuse with our experimentally estimated diffusion coefficient  $D_p$  along the long axis when polymeric ParA-ATP was not present either at any of the sites that the plasmid occupied or sites neighbouring the plasmid. In the presence of polymeric ParA-ATP, the plasmid was assumed to be tethered to such a polymer (via a ParB-*parC* complex), which prevented plasmid diffusion. At sites with a plasmid present, polymeric ParA-ATP could be converted into cytoplasmic ParA-ADP with reaction parameter  $k_B$ . Reflecting directed motion, at sites neighbouring a plasmid occupied by polymeric ParA-ATP, a plasmid could with reaction parameter  $k_{dm}$  move to the coordinate along the long axis of that ParA-ATP subunit, coinciding with conversion of that ParA-ATP into cytoplasmic ParA-ADP. For wild-type simulations, any plasmid in the system formed a hard wall to mobile ParA-ATP diffusion so that diffusing ParA-ATP molecules could not diffuse past a plasmid. See subsection 4.4.5 for details for further details on the simulation duration and model outputs.

We first adjusted the ParA-ATP polymerization rate to generate short filaments, of approximately 10 subunits in length (see Table 4.4 for parameters). Simulations again faithfully reproduced the equal spacing of plasmids along simulated growing nucleoids with varying numbers  $n_p$  of plasmids (Figure 4.6B,C,  $n_p = 1-4$ ) in good agreement with our experiments (Figure 4.1, Figure 4.2). By adjusting the ParA-ATP polymerization rate (Table 4.4), long continuous ParA polymer bundles could also be generated. In that case equal spacing could also be achieved (Figure 4.7). Intuitively, in both short and long filament cases, this occurs because in an irregularly spaced plasmid configuration, the unequal ParA concentrations on either side of a plasmid result in an unequal degree of ParA polymerization. This in turn



Reactions / Description	Propensities $p_t$
ParA-ATP diffusion $Am_{i,j} \rightarrow Am_{i+1,j}, i = 0..(L-2), j = 0..(S-1)$	$\frac{D_A}{dx^2} \cdot Am[i][j]$ if $P[i+1] = 0$ 0 otherwise
$Am_{i+1,j} \rightarrow Am_{i,j}, i = 0..(L-2), j = 0..(S-1)$	$\frac{D_A}{dx^2} \cdot Am[i+1][j]$ if $P[i] = 0$ 0 otherwise
$Am_{i,j} \rightarrow Am_{i,j+1}, i = 0..(L-1), j = 0..(S-2)$	$\frac{D_A}{dx^2} \cdot Am[i][j]$
$Am_{i,j+1} \rightarrow Am_{i,j}, i = 0..(L-1), j = 0..(S-2)$	$\frac{D_A}{dx^2} \cdot Am[i][j+1]$
Plasmid diffusion / immobilization $P_i \rightarrow P_{i+1}, i = 0..(L-2)$	$\frac{D_P}{dx^2} \cdot P[i]$ , if $\sum_{k=i,i+1,j=0..S-1} A[k][j]$ 0 otherwise
$P_{i+1} \rightarrow P_i, i = 0..(L-2)$	$\frac{D_P}{dx^2} \cdot P[i+1]$ , if $\sum_{k=i,i+1,i+2,j=0..S-1} A[k][j]$ 0 otherwise
Spontaneous ParA-ATP depolymerization $A_{i,j} \rightarrow Am_{i,j}, i = 0..(L-1), j = 0..(S-1)$	$k_{dp} \cdot A[i][j]$ if $P[i] = 0$ 0 otherwise
ParA-ATP polymerization $Am_{i,j} + Am_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i = 0..(L-2), j = 0..(S-1)$	$k_p \cdot Am[i][j] \cdot Am[i+1][j]$ if $P[i] = P[i+1] = A[i][j] = A[i+1][j] = 0$ 0 otherwise
$A_{i,j} + Am_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i = 0..(L-2), j = 0..(S-1)$	$k_p \cdot A[i][j] \cdot Am[i+1][j]$ if $P[i] = P[i+1] = A[i+1][j] = 0$ 0 otherwise
$Am_{i,j} + A_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i = 0..(L-2), j = 0..(S-1)$	$k_p \cdot Am[i][j] \cdot A[i+1][j]$ if $P[i] = P[i+1] = A[i][j] = 0$ 0 otherwise
ParA-ATP unbinding from plasmid $P_i + Am_{i,j} \rightarrow P_i + A_{ADP}, i = 0..(L-1), j = 0..(S-1)$	$k_{mB} \cdot P[i] \cdot Am[i][j]$
$P_i + Am_{i+1,j} \rightarrow P_i + A_{ADP}, i = 0..(L-2), j = 0..(S-1)$	$k_{mB} \cdot P[i] \cdot Am[i+1][j]$
$P_{i+1} + Am_{i,j} \rightarrow P_{i+1} + A_{ADP}, i = 0..(L-2), j = 0..(S-1)$	$k_{mB} \cdot P[i+1] \cdot Am[i][j]$
$P_i + A_{i,j} \rightarrow P_i + A_{ADP}, i = 0..(L-1), j = 0..(S-1)$	$k_B \cdot P[i] \cdot A[i][j]$
Directed plasmid motion $P_i + A_{i+1,j} \rightarrow P_{i+1} + A_{ADP}, i = 0..(L-2), j = 0..(S-1)$	$k_{dm} \cdot P[i] \cdot A[i+1][j]$
$P_{i+1} + A_{i,j} \rightarrow P_i + A_{ADP}, i = 0..(L-2), j = 0..(S-1)$	$k_{dm} \cdot P[i+1] \cdot A[i][j]$
Cytoplasmic ParA-ADP to ParA-ATP conversion $A_{ADP} \rightarrow A_{cyto}$	$k_W \cdot A_{ADP}$
ParA-ATP nucleoid binding $A_{cyto} \rightarrow Am_{i,j}, i = 0..(L-1), j = 0..(S-1)$	$\frac{k_{on}}{LS} \cdot ACYTO$

Table 4.3: Reactions and propensities used in the directed motion models.

results in an unequal amount of competitive directed motion events to each side, resulting in effective directed translocation over longer

length scales back towards an equally positioned state. Plasmid separation occurs when two nearby plasmids encounter two ParA-ATP structures extending in opposite directions away from the plasmids. The two ParA-ATP structures will then necessarily mediate a segregation event. The effect of directed movement in this model is clearest in the case of plasmid segregation events (Figure 4.6B, Figure 4.7A), where we see rapid segregation consistent with the fast segregation events observed experimentally (Figure 4.4A).

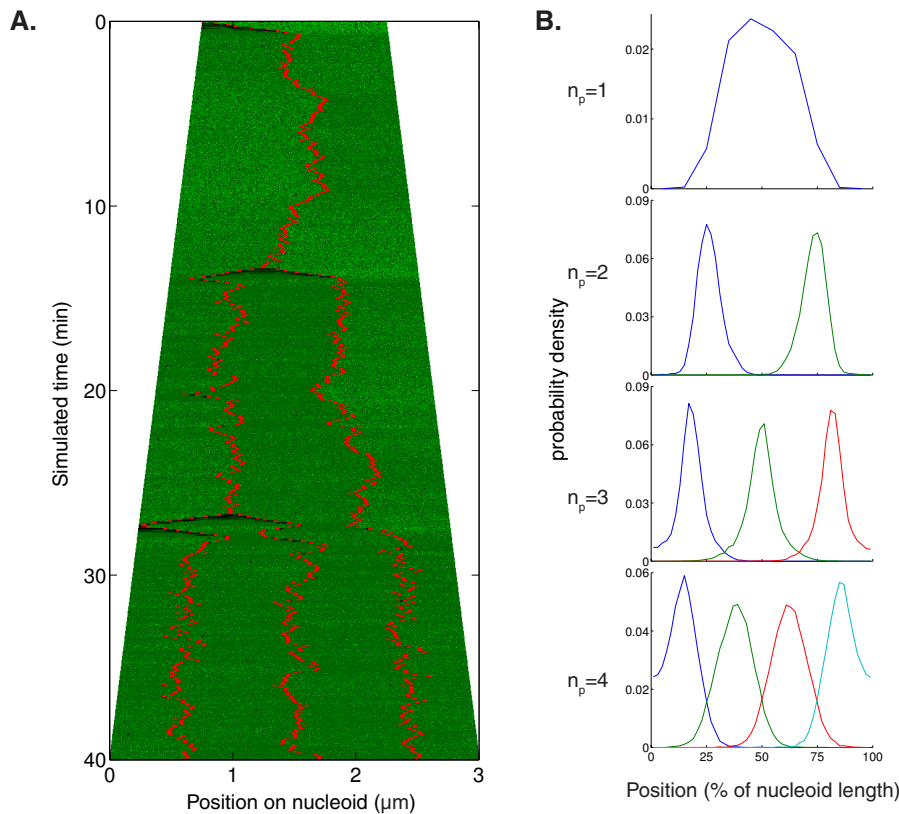


Figure 4.7: The directed motion model can equally space plasmids over the nucleoid, and is not critically dependent on the extent of ParA polymerization. (A) Typical simulation kymograph of the directed motion model with long polymers. Long polymers extend from nucleoid ends in a growing cell, where plasmid (red) is initially directed from a nucleoid edge to mid-cell by ParA (green) filament competition. After plasmid duplication, the system dynamically self-organizes to reacquire equal plasmid spacing. (B) Time-averaged plasmid position distributions for directed motion model with long polymers with  $n_p = 1 - 4$  plasmids. Simulated nucleoid growth and plasmid distributions obtained as in Figure 4.6C.

#### 4.2.6 *ParA-GFP oscillations are not continuously required for equal plasmid spacing.*

Intriguingly, simulations of the directed motion model did not generally produce sustained spatiotemporal oscillations of ParA across the nucleoid (short polymers: Figure 4.6B, long polymers: Figure 4.7A). A lack of sustained oscillations would therefore appear to be a common feature of models where competitive ParA structures generate equal plasmid spacing. This absence was unexpected, as prior experimental work had emphasized the oscillatory aspect of the ParA dynamics [156–158].

To experimentally test this key model prediction in an unbiased fashion, we experimentally measured the degree of ParA asymmetry in the *par2* system in a large dataset ( $n = 134$ ) of snapshots of ParA-GFP across the nucleoid. We summed 6 planes that are in focus from a Z-stack of ParA-GFP fluorescence signal images ( $dz = 0.2\mu m$ ), although the results are not different when using the ParA-GFP signal obtained from single confocal planes focused at mid-cell. Cell outlines, linear projections of ParA-GFP, *tetO*-TetR-mCherry and Hoechst stain fluorescence signal distributions, and *tetO*-TetR-mCherry foci positions were determined as described above for ParB-GFP. We confirmed that positioning of the *tetO*-TetR-mCherry foci from this dataset was similar to that measured previously [159]. We examined only cells with a single plasmid *tetO*-TetR-mCherry focus, where sustained oscillations should be easiest to infer. In those cells, the ParA-GFP fluorescence signal from pole to plasmid position was summed and divided by the respective pole-to-plasmid distance. This generates two ParA-GFP fluorescence densities  $I_L$  and  $I_R$  for either side extending to the two cell poles. This allows us to compute the normalized ParA asymmetry measure  $|I_L - I_R|/|I_L + I_R|$  (Figure 4.8A,B).

Irrespective of the plasmid position, a completely uniform fluorescence distribution would give an asymmetry value of zero. On the other hand, if all the ParA-GFP was located on one side of the plasmid the asymmetry measure would be one. Using a single confocal plane focused at mid-cell, we also computed the Hoechst asymmetry measure with respect to the plasmid position in the same manner.

Note that the ParA-GFP exposure time used here was 1.5s; clearly, we cannot measure asymmetries that occur on a timescale faster than this exposure time. However, the timescales of the plasmid and ParA-GFP dynamics are on the order of tens of seconds or longer and it is therefore unlikely that any significant asymmetry is being missed by our measurements.

As shown in [272] by using the same MT software package for analysis, the MinD-YFP asymmetry measure with respect to mid-cell follows an approximate sinusoidal oscillation over time, with a cell-length-dependent oscillation amplitude. In large cells the MinD-YFP oscillations are clearest with an amplitude  $|I_L - I_R|/|I_L + I_R|$  of around 0.6. To generate an asymmetry measure appropriate for the MinD-YFP oscillations, we sampled  $10^3$  time points  $t$  uniformly in  $[0, 2\pi]$  (which constitutes one period). We then computed for every time point  $|I_L - I_R|/|I_L + I_R| = |0.6 \cdot \sin(t)|$ . The resulting asymmetry distribution (Figure 4.8B) therefore reflects the experimental MinD-YFP asymmetry with respect to mid-cell in large cells [272]. In this way, we can directly compare the asymmetry present in the ParA-GFP and Hoechst signal distributions with that induced by the spatiotemporal oscillations of MinD-YFP. When we examined our whole distribution of cells exhibiting single plasmid *tetO*-TetR-mCherry foci, we found that the degree of ParA-GFP asymmetry (Figure 4.8A,B) was low in comparison with the well-established MinD spatiotemporal oscillator. Furthermore, the ParA-GFP asymmetry did not correlate with cell length (Figure 4.9A,  $R^2 = 0.08$ ), unlike the case of MinD-YFP [272].

We then generated asymmetry measures using our directed motion model. In simulation outcomes shown in Figure 4.6B (directed motion model with short polymers) and Figure 4.7A (directed motion model with long polymers), the plasmid position, cytoplasmic ParA-ADP, cytoplasmic ParA-ATP, nucleoid-bound mobile ParA-ATP and polymeric ParA-ATP levels on either side of the plasmid were output at regular time intervals of  $dt = 5s$  during a time period prior to plasmid duplication (first 2min and 1.5min of simulated time for directed motion model with short and long polymers respectively). Cy-

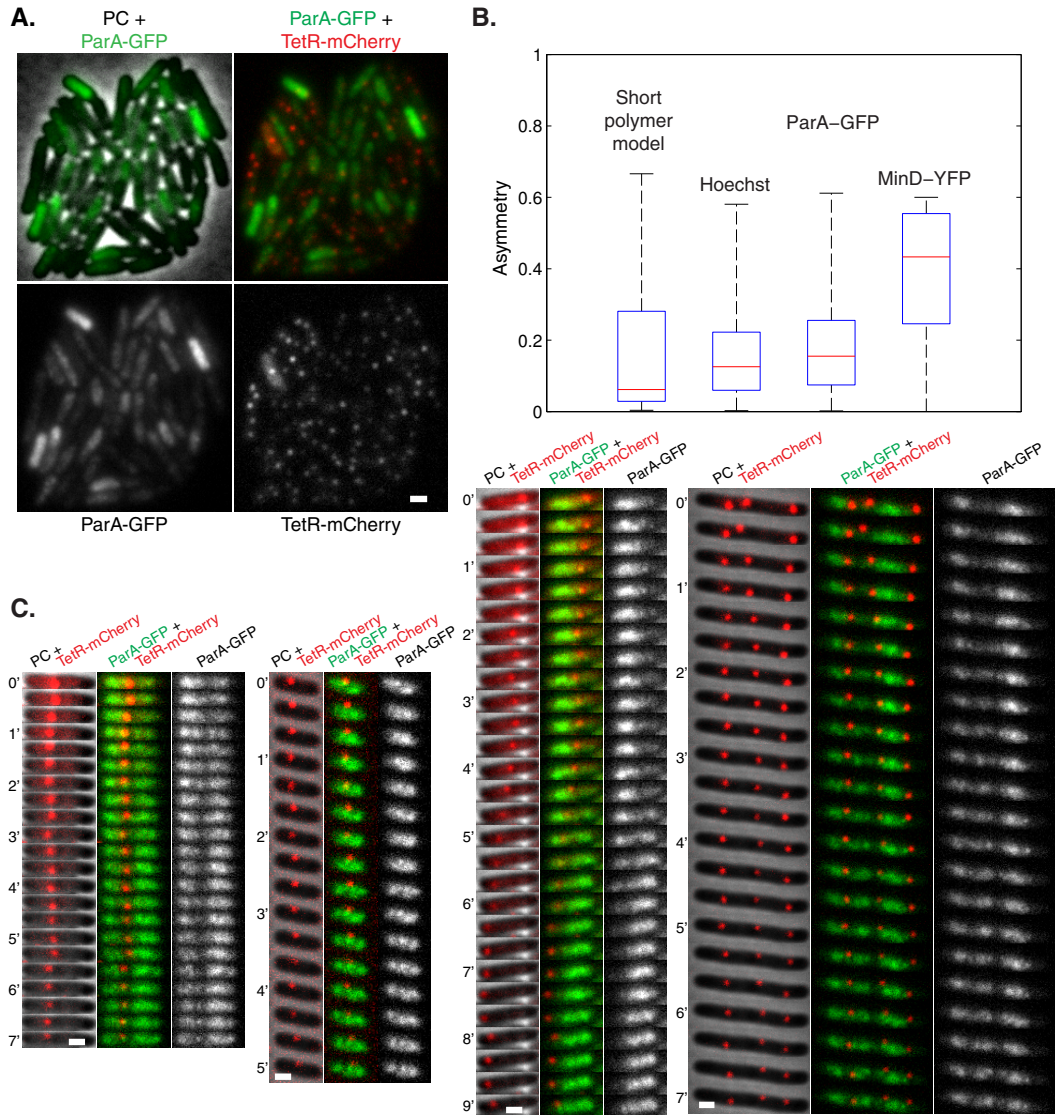


Figure 4.8: As predicted by the directed motion model, ParA-GFP distribution is relatively symmetric. (A) Localization of plasmids and summed Z-stack of ParA-GFP distributions in a field of cells. Scale bar: 1  $\mu$ m; plasmid: pSR233 (*mini-R1*, *par2*<sup>+</sup>, *P<sub>lac</sub>::parA::eGFP*, *tetO120*) in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). (B) ParA-GFP ( $n = 134$ ) asymmetry measure  $|I_L - I_R|/|I_L + I_R|$  using fluorescence densities  $I_L$ ,  $I_R$  on left, right sides of a plasmid focus along long cell axis. Comparison shown to the prediction of directed motion model with short polymers, Hoechst ( $n = 134$ ) and MinD-YFP case [272]. Box plots represent minimal, first quartile, median, third quartile and maximal values of asymmetries in all cases. (C) Time lapses showing localization of *par2*<sup>+</sup> pSR236 (*mini-R1*, *parC1*<sup>+</sup>, *parA*<sup>-</sup>, *parB*<sup>+</sup>, *parC2*<sup>+</sup>, *P<sub>lac</sub>::parA::eGFP*, *tetO120*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). Experiments performed by Florian Szardenings.

toplasmic ParA was assumed to be uniformly distributed throughout the cell (independently of the plasmid position), thus effectively only contributing to the denominator  $|I_L + I_R|$ . With this information we computed the ParA asymmetry using the same method as described for the experimental data. Results are shown in Figure 4.8B (short polymers) and Figure 4.9B (long polymers). It should be noted that according to both models, the ParA asymmetry remains very low once a plasmid is stably positioned at mid-cell, pushing the asymmetry distribution further towards zero over time. This is consistent with time lapses where stable equally spaced plasmid foci positioning correlates with ParA-GFP on either side of a plasmid focus (Figure 4.8C and [159]).

We also compared the ParA-GFP asymmetry to the Hoechst signal. This DNA stain labels the nucleoid itself, which is relatively uniform along the long cell axis [189–191]. Here, any asymmetry is not expected to depend on the plasmid foci positions. The Hoechst asymmetry distribution was indeed concentrated around relatively small values, but was apparently measurable within our approach (Figure 4.8B, Figure 4.9B). Importantly, we found that the ParA-GFP asymmetry measure had a similarly low value as for the Hoechst case (Figure 4.8B, Figure 4.9B, no significant difference, Kolmogorov-Smirnov test), and that for both the asymmetry is uncorrelated to the plasmid focus position (Figure 4.9C). We therefore conclude that for a single plasmid focus, ParA-GFP typically resides on both sides of a plasmid, with relatively little asymmetry or oscillation, as predicted by the directed motion model, irrespective of a weak (Figure 4.8B) or strong (Figure 4.9B) degree of polymerization.

Previous analyses had focused on plasmids migrating in the wake of retracting ParA-GFP structures [159]. Such events can transiently give rise to relatively high ParA-GFP asymmetries (see, for example, Figure 4.4A, Figure 4.8C). Accordingly, we conclude that ParA asymmetry or oscillations are not continuously required for *par2* mediated plasmid positioning. Transient asymmetry, including oscillations, instead likely arises from the dynamics needed to bring about equal plasmid spacing following a spatial perturbation or plasmid



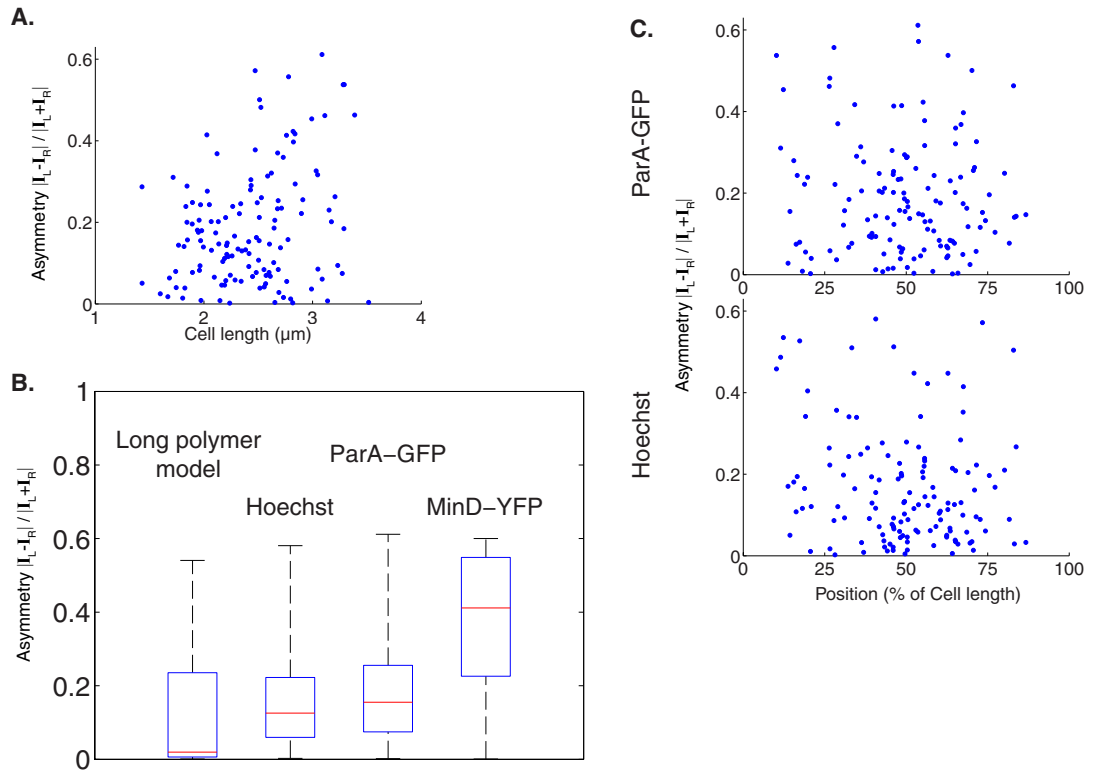


Figure 4.9: Hoechst DNA stain and ParA-GFP signal asymmetry are relatively low and uncorrelated to plasmid focus positioning. (A) Scatter plot of ParA-GFP asymmetry measure as a function of cell length ( $n = 134$ ). (B) ParA asymmetry prediction from the directed motion model with long polymers. Comparison shown to experimental ParA-GFP ( $n = 134$ ), Hoechst ( $n = 134$ ) and MinD-YFP distributions [272]. (C) Scatter plot of ParA-GFP and Hoechst asymmetry as a function of (a single) plasmid focus position relative to cell length. Experiments performed by Florian Szardenings.

duplication event (Figure 4.8C). Once the ParA distribution has returned to being relatively symmetric, this coincides with an equally spaced plasmid configuration (Figure 4.8C). Such dynamics can be seen in our model simulations (Figure 4.6B, Figure 4.7A): asymmetric during plasmid segregation events, but relatively symmetric otherwise. This analysis can therefore accommodate both our findings of a relatively symmetric ParA distribution with previous reports emphasizing asymmetry and oscillations. Overall, our finding of predominantly symmetric, non-oscillatory ParA dynamics may help to reconcile similar findings for ParA in other plasmid partitioning systems, such as for plasmid P1 [173, 183].

#### 4.2.7 *ParA-GFP forms structures within the nucleoid region.*

One required feature to achieve equal plasmid spacing is that the ParA-ATP should be organized in a 1d-like structure along the nucleoid as concluded above. However, it is unclear why ParA-ATP on either side of a plasmid would align in a coherent 1d-like structure with their ends coinciding with a plasmid. One potential explanation for this 1d-like behavior is that the ParA-ATP structures are sensitive to the overall nucleoid architecture. To test these features, we examined the localization of ParA-GFP and Hoechst signal simultaneously using optical sectioning in WT cells ( $n = 678$ ) without *par2*-carrying plasmids to prevent dynamic ParA-GFP structure disassembly. ParA-GFP intensity correlated well with the DNA stain (Figure 4.10A,B, Figure 4.11, Pearson's correlation coefficient  $r = 0.81$ ), indicating that ParA-GFP localization was indeed dependent on the underlying nucleoid. Importantly, by a Manders overlap coefficient analysis (see subsection 4.4.6 for details) we found that ParA-GFP overlaid more with Hoechst than the reverse (Figure 4.10C), indicating that ParA forms structures within the nucleoid region rather than uniformly covering the nucleoid. Although the resolution of our techniques does not allow identification of potential individual ParA polymers, in many cases we did observe extended 1d-like ParA-GFP structures on the nucleoid (Figure 4.10B, Figure 4.11). Care



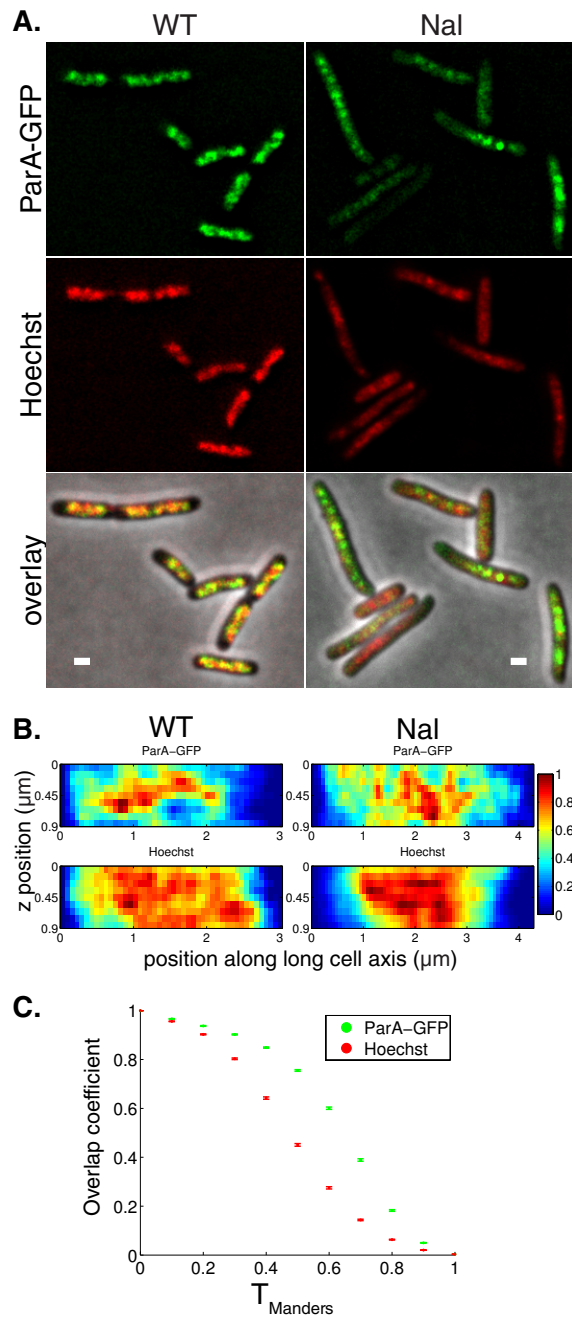


Figure 4.10: ParA forms structures within the nucleoid region. (A) Fluorescence localization of ParA-GFP (green), Hoechst DNA stain (red) and overlay, at mid-height through cell, taken from deconvolved Z-stacks showing structures that are disrupted with  $50\mu\text{g/ml}$  nalidixic acid treatment (Nal) compared to WT. Scale bar:  $1\mu\text{m}$ ; plasmid: pGE230 (mini-R1, *par*<sup>-</sup>, *P*<sub>lac</sub>::*parA*::eGFP). (B) Normalized fluorescence intensity profiles along the long cell axis for 9 in focus z heights ( $dz = 0.1\mu\text{m}$ ) resulting from deconvolved Z-stacks in representative WT and Nal-treated strains. (C) Manders overlap coefficients in WT cells (error bars: standard error of the mean,  $n = 678$ ) showing the fraction of ParA-GFP fluorescence intensity that overlaps with Hoechst DNA stain when the latter is above a threshold  $T_{\text{Manders}}$  (ParA-GFP, green) and the reverse (Hoechst, red). ParA-GFP overlaps more with Hoechst DNA stain (p-values ranging from  $10^{-12}$  to  $10^{-132}$ , see Materials and Methods) than the reverse. Experiments performed by Florian Szardenings.

must be taken in interpreting fluorescent localization studies due to potential artifacts, for example GFP-induced polymerization [275]. However, wild-type plasmid loss rates and plasmid foci positioning in cells expressing ParA-GFP argue against localization or polymerization artifacts in our case [156, 159].

#### 4.2.8 *Equal plasmid spacing is compromised in cells with a perturbed nucleoid.*

We reasoned that if ParA structures are reliant on the nucleoid morphology for their organization, then mutants/treatments that perturb the overall nucleoid structure should also exhibit alterations in ParA localization and therefore plasmid focus positioning (Figure 4.12A). We measured plasmid focus positioning in *mukE*, *mukF* and *matP* mutant strains, as well as in cells treated with the DNA gyrase inhibitor nalidixic acid (Nal), all of which exhibit defects in nucleoid organization [195, 196, 276]. Nucleoid length distributions were altered in all of these cases (Figure 4.12B) and, consistent with our hypothesis, there was in each case a similar deterioration in the fidelity of plasmid focus positioning ( $n_p = 1, 2$  in Figure 4.12C, Figure 4.13,  $n_p = 3, 4$  in Figure 4.13 and Figure 4.14) towards a random distribution (Figure 4.15). The latter was obtained as follows. For a given plasmid copy number  $n_p$  ranging from 1 to 4, we sampled a set of  $n_p$  independent plasmid positions from a one dimensional uniform distribution on  $[0, 100]$ . Each plasmid position set was ordered from small to large. We repeated this process  $10^5$  times and subsequently generated histograms of the ordered plasmid positions (Figure 4.15). Note that the experimental plasmid focus positioning histograms are significantly more ordered than these (compare Figure 4.1C and Figure 4.2B). Additionally, the expected interplasmid distance for randomly positioned plasmid pairs ( $n_p = 2$ ) is  $1/3$  of the nucleoid length, while the observed spacing is 0.5 (Figure 4.1D). These results demonstrate the effectiveness of the *parABC* plasmid positioning system. The deterioration in our nucleoid perturbed cells may not have been large enough to detect in stability assays [197, 198]. Similarly, in *E. coli*, *mukB* mutants perturbed plasmid po-

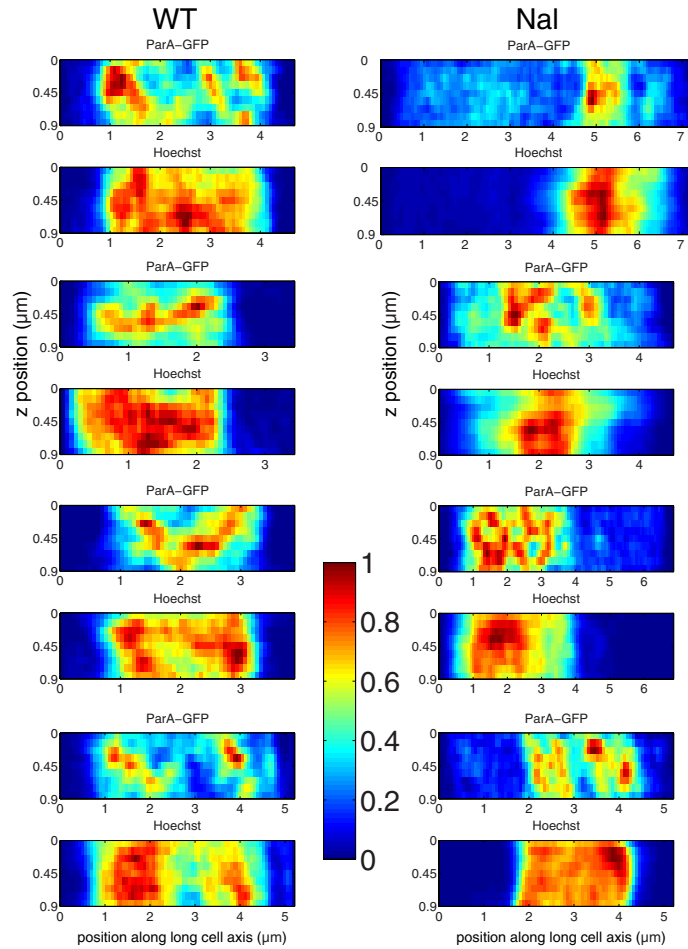


Figure 4.11: Correlation between Hoechst and ParA-GFP distributions. Normalized fluorescence intensity profiles along the long cell axis for 9 in focus  $z$  heights ( $dz = 0.1\mu\text{m}$ ) resulting from de-convolved Z-stacks in representative WT and Nal-treated strains. Many cases (representative examples shown) support the existence of linear ParA-GFP structures, although the inherent optical resolution of the imaging prohibits stronger conclusions about the presence or absence of narrow linear ParA-GFP filaments. For every cell having detectable Hoechst and ParA-GFP signals, the corresponding profiles were used for the systematic colocalization analyses. Experiments performed by Florian Szardenings.

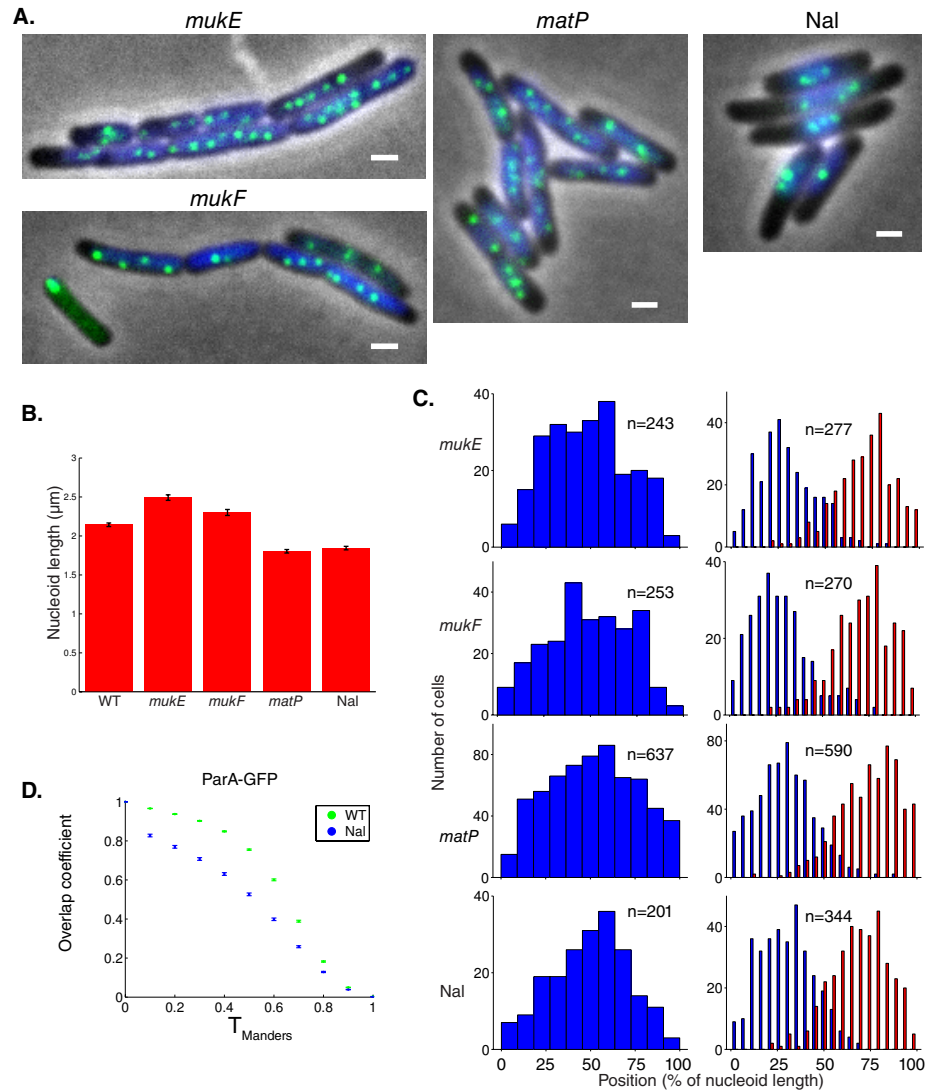


Figure 4.12: Nucleoid morphology disruption causes aberrant plasmid focus positioning. (A) Fluorescence localization of ParB-GFP (green) and Hoechst DNA stain (blue) in *mukE*, *mukF*, *matP* mutants and wild type cells treated with 50 µg/ml nalidixic acid (Nal). Scale bar: 1 µm; plasmid: pFS21 (mini-R1, *parC1*+, *parA*+, *parB::sfGFP*, *parC2*+) (B) Mean nucleoid length (error bars: standard error of the mean) of cells used for the plasmid positioning analysis shown in Figure 4.1B, Figure 4.2A and (C) in different strains: WT ( $n = 1695$ ), *mukE* ( $n = 1378$ ), *mukF* ( $n = 1555$ ), *matP* mutants ( $n = 2995$ ) and cells treated with nalidixic acid (Nal) ( $n = 1127$ ). According to unpaired t tests, all mutants and Nal show a mean differing from WT ( $p < 10^{-3}$ ). Although the average nucleoid length in *matP* mutants decreased, the average number of nucleoids per cell increased compared to WT ( $p < 10^{-41}$ ) due to a large fraction of cells exhibiting 2 nucleoids (using our half maximum criteria). This observation is consistent with the previously proposed function of *MatP* in preventing early segregation of duplicated Ter macrodomains. (C) Histograms of plasmid foci positions ( $n_p = 1, 2$ ) for mutants/treatments described in (A) relative to nucleoid size. According to Kolmogorov-Smirnov tests, all distributions are broader than WT (Figure 4.1C) with  $p < 10^{-2}$  except Nal  $n_p = 1$ :  $p < 0.05$ . (D) Manders overlap coefficients (error bars: standard error of the mean) of ParA-GFP comparing WT ( $n = 678$ ) and Nal-treated cells ( $n = 862$ ). Consistent with a decrease in the Pearson's correlation coefficient  $r$  ( $p < 10^{-38}$ ), ParA-GFP overlaps less with Hoechst in Nal-treated cells as compared to WT (p-values ranging from  $10^{-51}$  to  $10^{-144}$ ). Experiments performed by Florian Szardenings.

sitioning but without compromising plasmid stability, as found for the segregation mechanism mediated by ParM [277]. The deteriorations in plasmid positioning could have resulted from other effects, such as an induction of the SOS response in Nal-treated cells. However, the similarity of the altered plasmid positioning in all four cases instead suggests a common positioning defect based on nucleoid perturbation. This deterioration could also be due to an altered plasmid structure. However, at least for the case of *matP* we do not favor this hypothesis, due to the absence of *MatP* target sites (*matS*) on our test-plasmid.

To provide evidence that the above deterioration in plasmid posi-

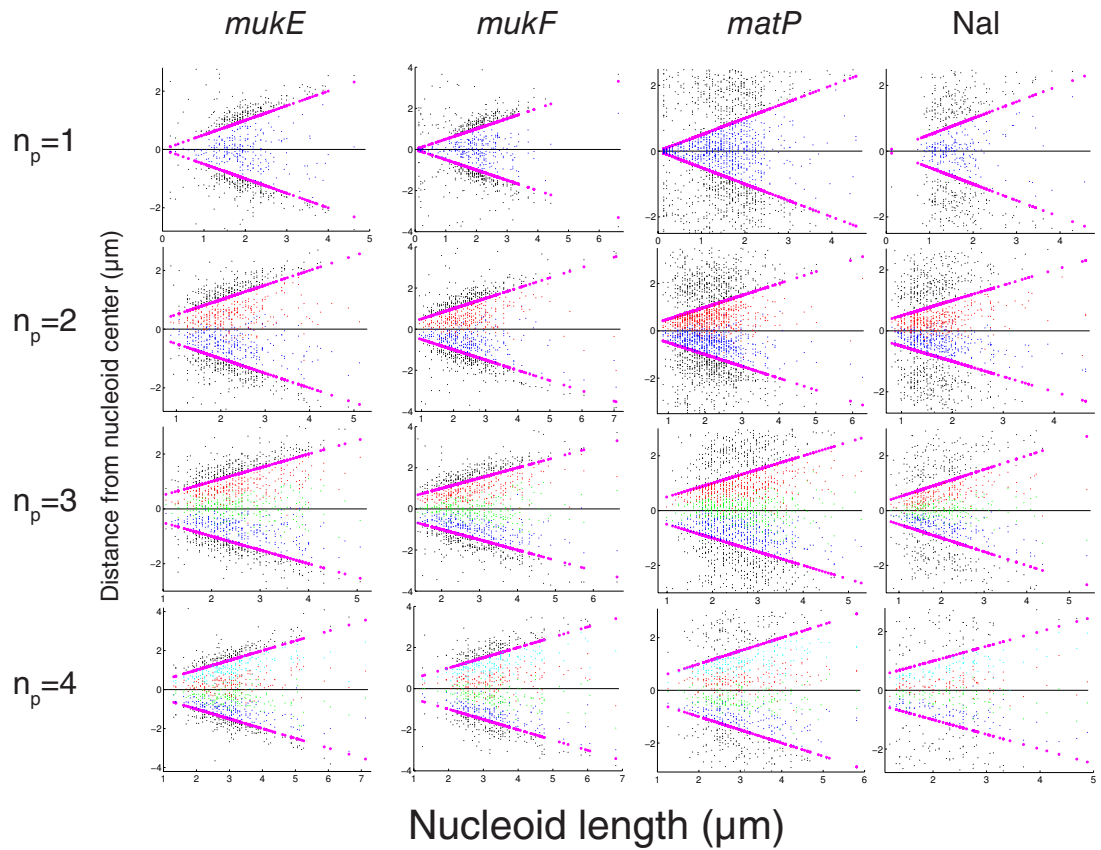


Figure 4.13: Scatter plot of  $n_p = 1 - 4$  plasmid foci positions (blue, green, red, cyan) with respect to nucleoid edges (purple) and cell edges (black) for *mukE*, *mukF* *matP* mutants and cells treated with  $50\mu\text{g/ml}$  nalidixic acid (Nal). Experiments performed by Florian Szardenings.

tioning arose from an altered ParA distribution, we systematically ex-

amined localization of ParA-GFP and Hoechst stain simultaneously in WT and Nal-treated cells ( $n = 862$ ). Visual comparison of the nucleoid shape between WT and nalidixic acid (Nal) treated cells (Figure 4.10A,B, Figure 4.11) revealed clear differences. In Nal-treated cells, the nucleoid signals, where present inside a cell, were more uniform along the long cell axis than in the WT (Figure 4.11 and Figure 4.16). Shape differences were also visible in the raw Z-stacks suggesting they were not artefacts of the deconvolution method. To quantify these shape differences in an unbiased and systematic manner, we performed the following analysis (Figure 4.16, for the preparatory image processing steps see subsection 4.4.6). We reasoned that a more uniform pattern would result in a profile along the long axis that resembled a first harmonic (first non-constant term of a Fourier expansion) between the nucleoid edges. Such a harmonic would not fit so well to a more spatially oscillating pattern that would arise, for example, from helical structures. We defined the maximal intensity value in the whole cell ( $I_{maxcell}$ ) and the maximal values at every  $z$  height ( $I_{max}(z)$ ). Using the Hoechst stain  $I_{maxcell}$  and the  $I_{max}(z)$  arising from the 9 relevant focus planes, we determined the half-maximum intensity locations along the long cell axis closest to the cell poles  $x_L$  and  $x_R$  at every  $z$  height. At every focus plane  $z$  height we could then define the 'first harmonic' function for  $x_L \leq x \leq x_R$ :

$$H(x, z) = \frac{I_{max}(z)}{I_{maxcell}} \left[ \frac{1}{2} + \frac{1}{2} \sin \left( \frac{\pi(x - x_L)}{x_R - x_L} \right) \right].$$

For every  $(x, z)$  we calculated the squared error  $SE(x, z)$  between the actual intensity value  $I(x, z)$  and  $H(x, z)$ :  $SE(x, z) = [I(x, z) - H(x, z)]^2$ . Lastly we summed over the SEs at every  $(x, z)$  and divided by the number of position points  $(x, z)$  to obtain a single measure of deviation  $SE_{cell}$  in a cell that is independent of the number of data points (and thus nucleoid size) and expression level variation between cells (because of normalization to  $I_{maxcell}$ ). We then performed a Wilcoxon rank sum test on the set of  $SE_{cell}$  comparing a population of WT cells with nucleoid-perturbed cells ( $n_{WT} = 678$  and  $n_{Nal} = 862$ ). Nucleoid shapes in Nal-treated cells were indeed altered ( $p < 10^{-149}$ ). This method did not detect a notable shape change in *matP* cells ( $n_{matP} = 579$ ), potentially due to our techniques not being sufficiently sensitive.

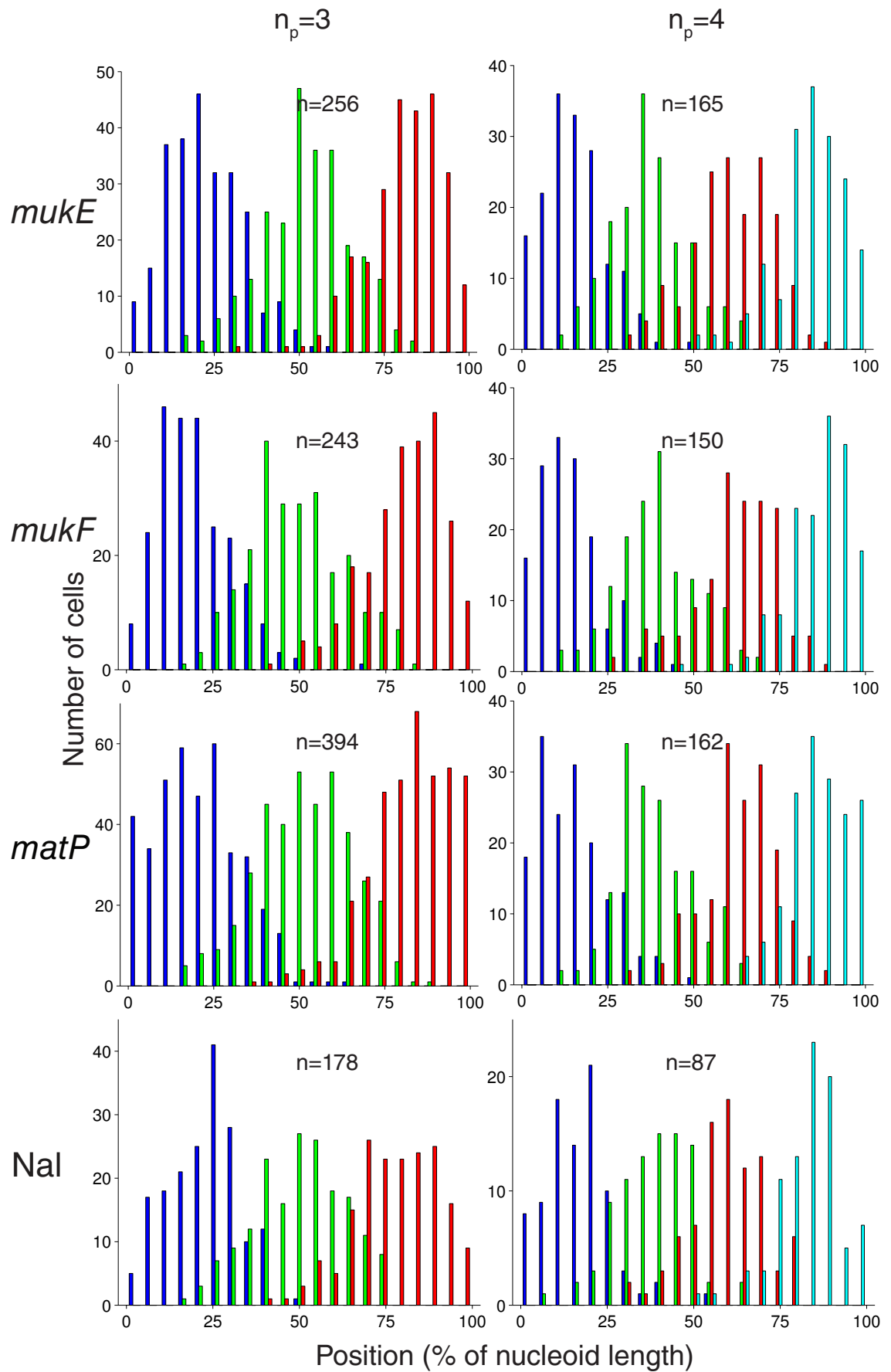


Figure 4.14: Histograms of  $n_p = 3, 4$  plasmid foci positions shown in Figure 4.13 relative to nucleoid size. Experiments performed by Florian Szardenings.

Moreover, visual inspection showed that the ParA-GFP distribution followed the nucleoid structure less closely in Nal treated cells than in WT (Figure 4.10A, Figure 4.11). To quantitate the colocalization of ParA-GFP and Hoechst signal in each cell, we also calculated, for every cell, the Pearson's correlation coefficient  $r$  using all the intensity values  $I_{\text{ParA-GFP}}(x, z)$  and  $I_{\text{Hoechst}}(x, z)$  [278]. Indeed the ParA-GFP distribution followed the nucleoid structure less closely because in Nal-treated cells the mean correlation coefficient ( $r = 0.68$ ) decreased from its WT value of 0.81 ( $p < 10^{-34}$ ). Note that in *matP* cells, however, we did not observe any significant alteration in intensity correlation ( $r = 0.80$  for *matP*).

We observed a decrease in the ParA-GFP overlap coefficient in Nal-treated as compared to WT cells (Figure 4.12D, see subsection 4.4.6 for details), which provided further evidence that the above deterioration in plasmid positioning arose from an altered ParA distribution. We did not observe any significant alteration in ParA-GFP overlap coefficient in *matP* as compared to the WT. This result was expected given that we could not detect any significant nucleoid structure alteration, as described above. Altogether, our findings support our hypothesis that the nucleoid provides a template for 1d-like ParA-ATP structure formation, which is partially compromised when the nucleoid structure is perturbed.

To reproduce this behavior in the directed motion model, we assumed that mobile DNA-bound ParA-ATP could now diffuse past a plasmid (see subsection 4.4.5 for details). This could be due to the disordered nucleoid structure resulting in a deteriorated ParA-ATP structure organization, thereby allowing ParA-ATP to spatially bypass ParB-*parC* complexes and compromise the ParA concentration differences between either side of a plasmid. The directed motion model with a weak or strong (Figure 4.17) degree of polymerization could then reproduce the observed plasmid focus distributions (Figure 4.12C).



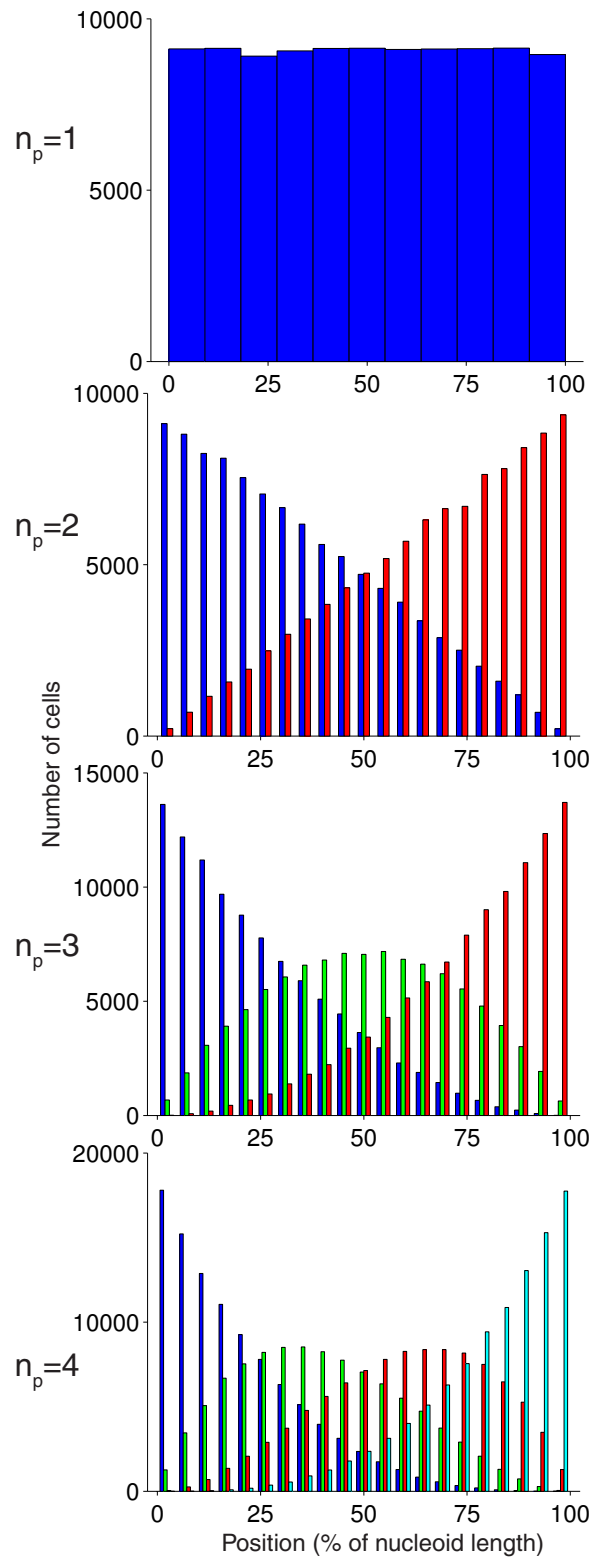


Figure 4.15: Histograms of  $10^5$  datasets for each of  $n_p = 1 - 4$ , where for each dataset plasmids are positioned in  $[0, 100]$  with a uniform distribution, independent from each other and consequently labeled  $1..n_p$  according to their position. This protocol induces an inherent spatial ordering. By comparing these distributions with the WT experimental data shown in Figure 4.1C ( $n_p = 1, 2$ ) and Figure 4.2B ( $n_p = 3, 4$ ) it is clear that the *parABC* system positions plasmid foci much more precisely, although the effect of active positioning becomes less clear as  $n_p$  increases.

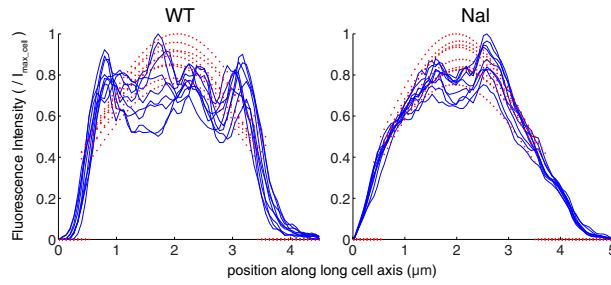


Figure 4.16: Correlation between Hoechst and ParA-GFP distributions. Graphical illustration of the unbiased systematic ‘first harmonic’ analysis of deconvolved 3d Hoechst signal inside representative cells in WT and Nal-treated strains. The Hoechst (blue) profiles indicate the signal intensities (integrated over the cell width) along the long cell axis at 9 in focus  $z$  heights with corresponding ‘first harmonics’ (dotted red curves). Fluorescence signal distributions deviate significantly more from the first harmonics in WT compared to Nal-treated cells (Wilcoxon rank sum test,  $p < 10^{-149}$ ), showing that Hoechst DNA stain distributions are perturbed in the latter. This analysis is independent of nucleoid length, which is altered in Nal-treated strains as compared to WT (Figure 4.12B). Experiments performed by Florian Szardenings.

### 4.3 Discussion

Stable DNA inheritance is important for the viability of essentially all organisms. In bacteria, the *parABC* genes have a major role in this process for plasmid DNA [155]. In this study, we have investigated how *E. coli* utilizes the *par2* partitioning system from plasmid pB171. We have for the first time provided a robust mechanistic explanation for how plasmids are equally spaced over the nucleoid, a process vital for the fidelity of low copy number plasmid inheritance. We propose that competing ParA structures function to direct plasmid movement over the nucleoid to equally spaced positions. This mechanism is likely relevant to other *parABC* systems that move and position sub-cellular objects.

It has previously been proposed that plasmid positioning is controlled by concentration gradients of ParA-ATP over the nucleoid, caused by plasmid-associated ParB-*parC* complexes mediating ParA-ATP hydrolysis [166, 173, 176, 177, 180, 186]. In this so-called diffusion-ratchet mechanism [166, 176, 177, 180], it has remained unclear whether such a mechanism could actually mediate equal plasmid spacing, and

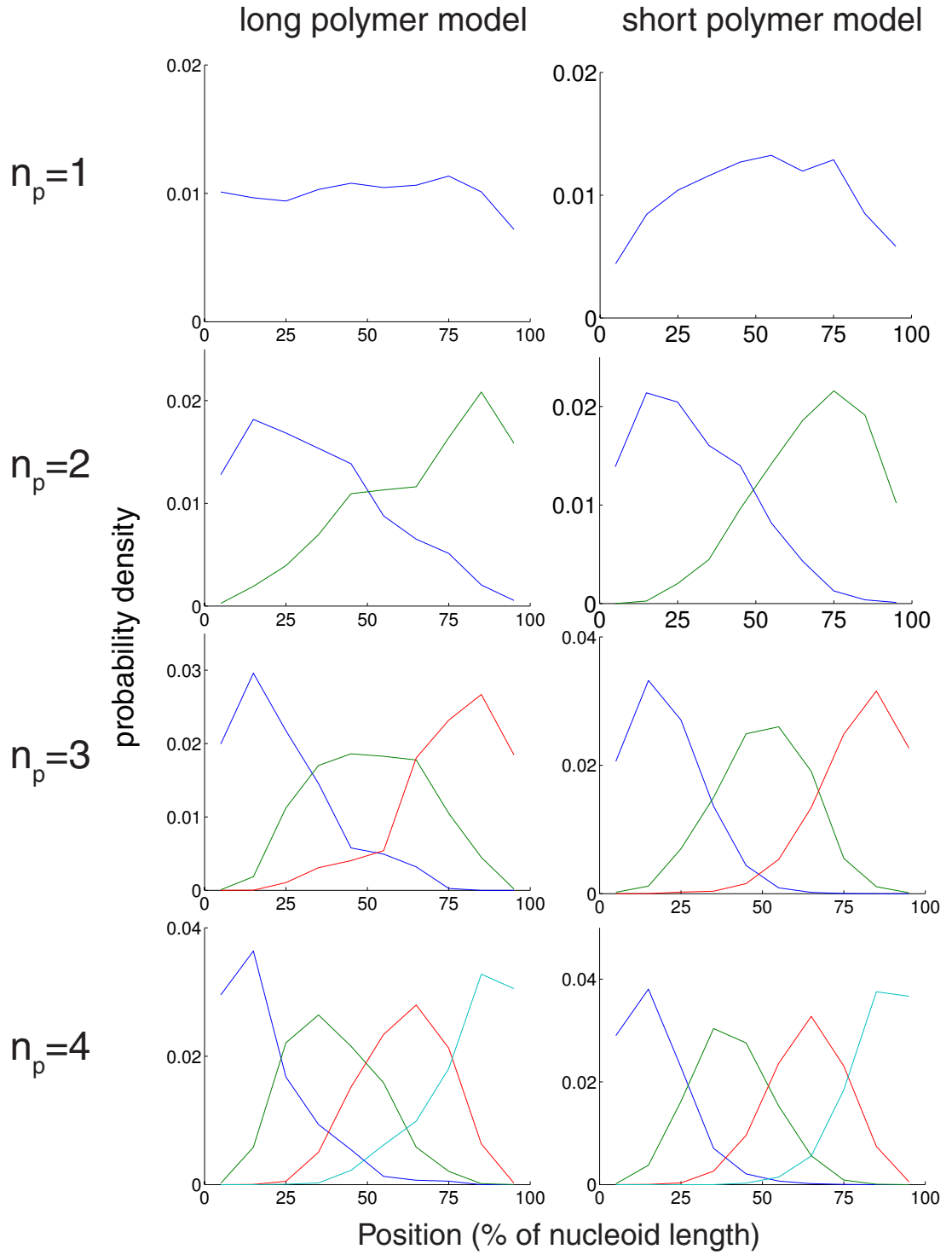


Figure 4.17: Time-averaged plasmid position distributions for directed motion models with short and long polymers for  $n_p = 1 - 4$  on simulated growing nucleoids without plasmid duplication. Results obtained from 124 independent simulations, where ParA-ATP could now diffuse past a plasmid (see subsection 4.4.5 for details).

if so, which specific properties of the system were key. In particular, it was left unclear how ParA actually influenced plasmid movement [166, 177, 180], e.g. through immobilizing plasmids or actively directing their motion through a chemophoresis force [176, 186]. Furthermore, although the diffusion-ratchet mechanism did not strictly preclude some degree of ParA polymerization, its gradient-aspect was emphasized as opposed to polymerization [166, 173, 176, 177, 180], leaving open the potential importance of polymerization. To provide elucidation of these key issues, we have therefore performed a mathematical analysis, which has led to predictions that we have experimentally verified.

We found that ParA-ATP nucleoid-binding, followed by diffusion over the nucleoid, and subsequent ParB-*parC*-stimulated ParA unbinding in a 1d model, is sufficient to generate dynamic ParA-ATP concentration gradients on either side of a plasmid. We have further shown that these ParA concentrations on either side of a plasmid are only symmetric in the case of equally spaced plasmids; unequally-spaced plasmid configurations will cause the ParA gradient to be steeper on one side rather than the other. Fundamentally, this asymmetry arises from two key properties: (i) a greater space for binding of ParA on one side as opposed to the other in unequally-spaced configurations, and (ii) ParA only being returned to the cytoplasm at discrete plasmid positions occupied by ParB-*parC*. The combination of these two features leads to the ParA density being increased in larger versus smaller inter-plasmid regions and hence to asymmetric ParA concentrations in unequally spaced plasmid configurations. According to our analysis, all that is then required for equal plasmid spacing is that the plasmids have a means to preferentially move up the locally steepest ParA concentration gradient and thus locate the equally spaced configuration with symmetric, competitive ParA concentrations around each plasmid. The exact means of plasmid translocation is therefore not critical; all that is important is that such movement can occur.

With this general framework established, we then investigated which specific means of plasmid movement up a concentration gradient

were possible, and which was implemented for the *par2* segregation system. We first developed a diffusion/immobilization model and found that such a model could indeed lead to plasmid movement up a ParA gradient, as the plasmid tends to become trapped in regions of higher ParA concentration. However, when we tested this model experimentally, its predictions did not verify: in particular, plasmid mobility was higher in the presence rather than the absence of *par2*, and overall free plasmid mobility was too low to allow the experimentally-observed rapid plasmid segregation following duplication events. This intrinsically low mobility agrees with earlier measurements [161, 199, 200] and is likely a general feature for relatively large intracellular components, given the glass-like properties of the cytoplasm [199].

We then considered active means of ParA-mediated plasmid movement. In particular, we assumed that ParA-ATP could form polymeric filaments, which could subsequently depolymerize through the action of plasmid-associated ParB-*parC*. In this case, ParA-ATP could bind to the nucleoid, diffuse and then subsequently polymerize to form ParA polymers, with the degree of polymerization influenced by the overall ParA concentration at a particular location. We found that ParA polymer models could naturally explain enhanced plasmid mobility in the presence of *par2*, as well as rapid plasmid segregation events, much more satisfactorily than the diffusion-immobilization model, regardless of whether long or short ParA polymers were formed. This finding in particular shows that our directed motion model is sufficiently general to explain equal plasmid spacing as found in various *parABC* systems with different extents of ParA polymerization [159, 166, 180]. In addition, we note that this mechanism does not critically depend on ParA-ATP binding to the nucleoid as a dimer. A scenario where ParA polymerizes to a certain extent cytoplasmically, and subsequently binds and diffuses on the nucleoid before polymerizing further into immobile filaments, could also suffice.

A key aspect of our models is competition between ParA structures on either side of a plasmid to direct plasmid movement. Therefore our model predicts a comparatively symmetric ParA distribution on

average, a prediction which we experimentally verified. We note here that such competition makes the system dynamics robust to alterations in ParA expression levels, since it is only the relative rather than absolute ParA levels on either side of a plasmid that are critical. This analysis potentially explains why cells with variable amounts of ParA-GFP (Figure 4.18A), still possess functional segregation systems with low plasmid loss rates [159].

In the above polymer models, the movement of a plasmid is assumed to be directed by retracting ParA structures. The precise nature of this short-ranged directed motion is not specified by our analysis, and could include locally biased plasmid diffusion along a retracting polymer in a 'burnt-bridge' mechanism [167] or even direct pulling [185]. This arbitrariness is a special case of our more general result that the mechanism by which a plasmid is able to move up a ParA concentration gradient is not important, only that such movement is possible. Other mechanisms of directed motion are also plausible. One possibility is that ParA-ATP does not polymerize at all, but nevertheless forms dense structures on the nucleoid with many ParA-ATP contacting a plasmid at any given time. In this variant, biased diffusion through an analog of a 'burnt-bridge' mechanism is still possible. Another possibility is a DNA-relay, where directed motion is generated by the elastic dynamics of the nucleoid DNA to which ParA-ATP dimers are bound [169]. Moreover, plasmid diffusion seems not always required for directed plasmid movement. Brownian dynamics simulations based on ParB-*parC*-mediated disassembling ParA polymer bundles can both tether and pull plasmids simultaneously without the need for plasmid diffusion [185]. We propose that distinct underlying translocation mechanisms, as exemplified above, could be responsible for directed motion in different *parABC* systems and yet still attain similar equal plasmid spacing.

For our models to generate equal plasmid spacing, ParA should be organized into a 1d-like configuration along the nucleoid. If ParA were not organized in this way, it would be possible for ParA to diffuse around the sides of a plasmid without encountering the hydrolyzing effect of the ParB-*parC* complex. This would equalize the

ParA concentrations on both sides even in the case of asymmetrically placed plasmids, leading to failure of the equal spacing mechanism. Potentially such ParA structures could consist of long ParA polymer bundles, or an extended region containing short ParA polymers or dimers. Importantly, in this work, we have provided experimental evidence for such ParA structure formation within the nucleoid region. Interestingly, it has been reported that the *E. coli* chromosome adopts a helical shape [189, 190]. Potentially the ParA structures could be preferentially located within a 'valley' in this configuration, thereby naturally generating a 1d-like appearance, even for dimers or short polymers. Consistent with these concepts, we found experimentally that plasmid positioning is compromised in nucleoid perturbed strains. ParA structures could also provide a high enough ParA concentration to ensure plasmid tethering and directed plasmid motion, whilst preventing plasmids from diffusing away from the nucleoid, a process which would compromise regular positioning. Further investigation of the exact involvement of the nucleoid in intracellular cargo positioning is therefore an important future goal.

## 4.4 Materials and Methods

### 4.4.1 *Mathematical derivation that dynamic ParA concentrations can generate equal plasmid spacing.*

#### EXPRESSION FOR $A(x)$

In an earlier section we described how ParA symmetry results in equal plasmid spacing. There we assumed that the ParA concentration at each plasmid is zero in order to gain understanding of the mechanism and keep the analysis concise. Here we show that equal plasmid spacing can also be achieved in cases where the ParA concentration at each plasmid is not zero, a more realistic scenario. In fact all that is required is the less stringent condition that ParB-*parC* complexes mediate ParA turnover at the plasmid.

We recall that we model the nucleoid as a 1d system of length  $L$  (along the cell long axis) on which ParA-ATP and plasmids can interact.  $A(x, t)$  denotes the nucleoid-associated ParA-ATP concentration at position  $x$  relative to one nucleoid edge at time  $t$ . Let  $x_1(t) \dots x_{n_p}(t)$  be

the positions of the  $n_p$  plasmids. At each plasmid, ParA-ATP can be hydrolyzed with rate  $k_B$ , turning Par-ATP into a cytoplasmic ParA form, with copy number  $A_c(t)$ . After sufficiently long timescales, the cytoplasmic ParA becomes competent to bind the nucleoid again, with rate  $J(A_c(t))$ . Once bound to the nucleoid, ParA-ATP can diffuse along the nucleoid with diffusion constant  $D$ . This system can be described by the deterministic reaction-diffusion equations:

$$\begin{aligned} \frac{\partial A(x, t)}{\partial t} &= D \frac{\partial^2 A(x, t)}{\partial x^2} - k_B \sum_{i=1}^{n_p} A(x_i(t)) \delta(x - x_i(t)) + \frac{J(A_c(t))}{L} \\ \frac{dA_c(t)}{dt} &= k_B \sum_{i=1}^{n_p} A(x_i(t)) - J(A_c(t)) \\ \text{Boundary Conditions : } \frac{\partial A(x, t)}{\partial x} \Big|_{x=0} &= 0 = \frac{\partial A(x, t)}{\partial x} \Big|_{x=L} \text{ for all } t. \end{aligned} \quad (4.4)$$

Here  $\delta(x)$  indicates the Dirac delta function. In line with the experimental evidence we assume that cytoplasmic ParA binds to the nucleoid with a slow rate  $k_W$  after becoming cytoplasmic:  $J(A_c(t)) = k_W A_c(t)$ . Assuming a given (time-independent) total ParA copy number  $A_T$  in the system, we will find that  $J$  is determined implicitly:  $A_T = A_c(t) + \int_0^L A(J(A_c(t)), x, t) dx$ . We will use this relation to calculate the steady state  $\left(\frac{\partial A(x, t)}{\partial t} = 0 = \frac{dA_c(t)}{dt}\right)$  solution  $A(x)$  for any given plasmid configuration  $x_1 \dots x_{n_p}$  with plasmid copy number  $n_p = 1, 2$  in terms of  $A_T$  and the other parameters  $L, D, k_B$  and  $k_W$ . Note that the dimensions of  $k_B$  are length/time due to the dimensionality of the Dirac delta function  $\delta(x)$ . The procedure described here generalizes to any  $n_p$ .

Recall from the main text that by invoking a separation of timescales between ParA diffusion and plasmid motion, Equation 4.4 reduces in steady state to:

$$\begin{aligned} D \frac{d^2 A(x)}{dx^2} &= k_B \sum_{i=1}^{n_p} A(x_i) \delta(x - x_i) - \frac{J}{L} \\ \text{Boundary Conditions : } \frac{dA(x)}{dx} \Big|_{x=0} &= 0 = \frac{dA(x)}{dx} \Big|_{x=L} \\ k_W A_c &= J = k_B \sum_{i=1}^{n_p} A(x_i) \text{ (flux balance.)} \end{aligned} \quad (4.5)$$



Equation 4.5 can be solved for  $A(x)$  in terms of  $J$  and  $A(x_1) \dots A(x_{n_p})$  by using the Neumann boundary conditions:

$$A(x) = \begin{cases} \frac{J}{2LD} (x_1^2 - x^2) + A(x_1) & \text{if } 0 \leq x \leq x_1, \\ \frac{k_B}{D} \sum_{i=1}^j A(x_i) (x - x_j) + \frac{J}{2LD} (x_j^2 - x^2) + A(x_j) & \text{if } x_j \leq x \leq x_{j+1}, 1 \leq j < n_p, \\ \frac{J}{2LD} ((L - x_{n_p})^2 - (L - x)^2) + A(x_{n_p}) & \text{if } x_{n_p} \leq x \leq L. \end{cases} \quad (4.6)$$

Note that in order to be physically relevant we assume our solution  $A(x)$  to be a continuous function of  $x$ . This generates the following recursive relations between the concentrations at the plasmids:

$$A(x_j) = \frac{k_B}{D} \sum_{i=1}^{j-1} A(x_i) (x_j - x_{j-1}) + \frac{J}{2LD} (x_{j-1}^2 - x_j^2) + A(x_{j-1}) \text{ for } 1 < j \leq n_p. \quad (4.7)$$

Now we focus on obtaining  $A(x_1)$  in the case  $n_p = 1$  for any plasmid position  $x_1$ . Integration over the nucleoid results in  $\int_0^L A(x) dx = A(x_1)L + \frac{J}{D} \left[ \frac{L^2}{3} - x_1L + x_1^2 \right]$ . From conservation of total ParA in the system,  $J$  is determined implicitly in terms of  $A(x_1)$ :  $J = k_W A_c = k_W \left( A_T - \left[ A(x_1)L + \frac{J}{D} \left[ \frac{L^2}{3} - x_1L + x_1^2 \right] \right] \right)$ . Solving for  $J$  results in:  $J = \frac{A_T - A(x_1)L}{\frac{1}{k_W} + \frac{1}{D} \left[ \frac{L^2}{3} - x_1L + x_1^2 \right]}$ . Now we use the flux balance constraint in Equation 4.5 to obtain the desired  $A(x_1)$ :

$$\text{for } n_p = 1: A(x_1) = \frac{A_T}{L + \frac{k_B}{k_W} + \frac{k_B}{D} \left[ \frac{L^2}{3} - x_1L + x_1^2 \right]}.$$

Now that  $J$  and  $A(x_1)$  are obtained, the ParA distribution on the nucleoid as described by Equation 4.6 is fully determined. For  $n_p > 1$  this procedure can be repeated to find  $A(x_1)$ . Due to the continuity constraint, Equation 4.7, all the other  $A(x_j)$  are then also determined. We illustrate this for  $n_p = 2$ : piecewise integration over the nucleoid results in:  $\int_0^L A(x) dx = A(x_1)L + \frac{J}{D} \left( \frac{2}{3}L^2 + \frac{3}{2}x_2^2 - 2Lx_2 - \frac{1}{3L}x_2^3 + \frac{1}{2}x_1^2 \right) + \frac{k_B}{2D} A(x_1) (x_2 - x_1) (2L - x_1 - x_2)$ . Now we use the conservation of total ParA again to obtain  $J$  in terms of  $A(x_1)$ :  $J = \frac{A_T - A(x_1)L - \frac{k_B}{2D} A(x_1) (x_2 - x_1) (2L - x_1 - x_2)}{\frac{1}{k_W} + \frac{1}{D} \left[ \frac{2}{3}L^2 + \frac{3}{2}x_2^2 - 2Lx_2 - \frac{1}{3L}x_2^3 + \frac{1}{2}x_1^2 \right]}$ . Again turning to the flux balance condition, whilst realizing that by Equation 4.7  $A(x_2)$  is known in terms of  $A(x_1)$  and  $J$ , we find:  $J = k_B \left[ A(x_1) + \frac{k_B}{D} A(x_1) (x_2 - x_1) + \frac{J}{2LD} (x_1^2 - x_2^2) + A(x_1) \right]$ . Now we solve for  $A(x_1)$ :

$$A(x_1) = \frac{A_T}{L + \frac{k_B}{2D} (x_2 - x_1) (2L - x_1 - x_2) + \left[ 2k_B + \frac{k_B^2}{D} (x_2 - x_1) \right] B},$$

$$\text{with } B = \frac{\frac{1}{k_W} + \frac{1}{D} \left[ \frac{2}{3} L^2 + \frac{3}{2} x_2^2 - 2Lx_2 - \frac{1}{3L} x_2^3 + \frac{1}{2} x_1^2 \right]}{1 + \frac{k_B}{2LD} (x_2^2 - x_1^2)}.$$

This determines  $A(x)$  for  $n_p = 2$  completely.

#### DERIVATION THAT PARA SYMMETRY IMPLIES EQUAL PLASMID SPACING

In this section we derive that for any  $n_p \geq 1$  for our steady state solution  $A(x)$ , the following statement holds:

for all  $j$  and  $x$  such that  $x_{j-1} \leq x_j - x \leq x_j$  and  $x_j \leq x_j + x \leq x_{j+1}$  :

$$A(x_j - x) = A(x_j + x) \implies \text{for all } j : x_j = \frac{L}{2n_p} + \frac{L}{n_p} (j - 1)$$

Here it is understood that  $1 \leq j \leq n_p$ , indicating the label of the  $j^{\text{th}}$  plasmid that are assumed to be ordered (without loss of generality):  $x_1 \leq \dots \leq x_{n_p}$ . Furthermore we define  $x_0 = 0$  and  $x_{n_p+1} = L$ . To show this we use the expressions for  $A(x)$  described in Equation 4.6. First we focus on  $j = n_p$ . Let  $x_{n_p-1} \leq x_{n_p} - x \leq x_{n_p}$  and  $x_{n_p} \leq x_{n_p} + x \leq L$  and  $A(x_{n_p} - x) = A(x_{n_p} + x)$ :

$$\begin{aligned} \frac{k_B}{D} \sum_{i=1}^{n_p-1} A(x_i) ((x_{n_p} - x) - x_{n_p-1}) + \frac{J}{2LD} (x_{n_p-1}^2 - (x_{n_p} - x)^2) + A(x_{n_p-1}) = \\ \frac{J}{2LD} ((L - x_{n_p})^2 - (L - (x_{n_p} + x))^2) + A(x_{n_p}) \implies \\ \left( \frac{k_B}{D} \sum_{i=1}^{n_p-1} A(x_i) + \frac{J}{D} \right) x = \frac{2J}{LD} x_{n_p} x \text{ for all } x \implies \\ x_{n_p} = \frac{L}{2} \left( 1 + \frac{k_B}{J} \sum_{i=1}^{n_p-1} A(x_i) \right). \end{aligned} \quad (4.8)$$

Note that we used the continuity requirement (Equation 4.7) here for  $A(x_{n_p})$  on the right hand side. In the special case of  $n_p = 1$ , the left hand side of the equation can be replaced by  $\frac{J}{2LD} (x_1^2 - (x_1 - x)^2) + A(x_1)$ , which using the same procedure leads straightforwardly to the desired result  $x_1 = \frac{L}{2}$ . Now in the case of  $n_p > 1$ , we proceed with  $j = 1$ . Let  $0 \leq x_1 - x \leq x_1$  and  $x_1 \leq x_1 + x \leq x_2$  and  $A(x_1 - x) = A(x_1 + x)$ :

$$\begin{aligned} \frac{J}{2LD} (x_1^2 - (x_1 - x)^2) + A(x_1) = \\ \frac{k_B}{D} A(x_1) ((x_1 + x) - x_1) + \frac{J}{2LD} (x_1^2 - (x_1 + x)^2) + A(x_1) \implies \\ 2 \frac{J}{L} x_1 x = k_B A(x_1) x \text{ for all } x \implies \\ x_1 = \frac{L k_B}{2J} A(x_1). \end{aligned}$$

Proceeding with  $1 < j < n_p$ , we let  $x_{j-1} \leq x_j - x \leq x_j$  and  $x_j \leq x_j + x \leq x_{j+1}$  and use again Equation 4.7 to replace  $A(x_j)$ :

$$\begin{aligned} \frac{k_B}{D} \sum_{i=1}^{j-1} A(x_i) ((x_j - x) - x_{j-1}) + \frac{J}{2LD} (x_{j-1}^2 - (x_j - x)^2) + A(x_{j-1}) = \\ \frac{k_B}{D} \sum_{i=1}^j A(x_i) ((x_j + x) - x_j) + \frac{J}{2LD} (x_j^2 - (x_j + x)^2) + A(x_j) \Rightarrow \\ \frac{J}{2L} (4x_j x) = k_B \left[ 2 \sum_{i=1}^{j-1} A(x_i) + A(x_j) \right] x \text{ for all } x \Rightarrow \\ x_j = \frac{Lk_B}{J} \sum_{i=1}^{j-1} A(x_i) + \frac{Lk_B}{2J} A(x_j). \end{aligned}$$

Now we have direct relations between the concentrations at each plasmid and the position of the plasmids for symmetric concentrations. Note that by mathematical induction, it is straightforward to show that for  $1 \leq j < n_p$ :

$$A(x_j) = \frac{2J}{Lk_B} \left( x_j + 2 \sum_{i=1}^{j-1} (-1)^{j-i} x_i \right). \quad (4.9)$$

Now we define the plasmid spacings  $z_j := x_j - x_{j-1}$ . Note that by Equation 4.9 for  $1 < j < n_p$ :  $z_j = \frac{Lk_B}{2J} (A(x_j) + A(x_{j-1}))$ . We can then replace all  $A(x_i)$  in Equation 4.7 in terms of the  $x_j$  and  $z_j$  and subsequently solve for the spacing and positions. First we rewrite Equation 4.7 as:

$$\begin{aligned} A(x_j) + A(x_{j-1}) &= \frac{k_B}{D} \sum_{i=1}^{j-1} A(x_i) z_j - \frac{J}{2LD} z_j (z_j + 2x_{j-1}) + 2A(x_{j-1}) \Rightarrow \\ \frac{2}{k_B} z_j &= \begin{cases} \frac{2}{D} \left( \sum_{m=1}^{\frac{j}{2}-1} z_{2m+1} + x_1 \right) z_j - \frac{1}{2D} z_j (z_j + 2x_{j-1}) + \frac{4}{k_B} \left( x_{j-1} + 2 \sum_{i=1}^{j-2} (-1)^{j-1-i} x_i \right) & \text{if } j \text{ even} \\ \frac{2}{D} \sum_{m=1}^{\frac{j-1}{2}} z_{2m} z_j - \frac{1}{2D} z_j (z_j + 2x_{j-1}) + \frac{4}{k_B} \left( x_{j-1} + 2 \sum_{i=1}^{j-2} (-1)^{j-1-i} x_i \right) & \text{if } j \text{ odd} \end{cases} \end{aligned}$$

These equations are essentially quadratic equations in  $z_j$ . Using mathematical induction we will now show that  $z_j = 2x_1$  for all  $2 \leq j < n_p$ . We start with the base case  $j = 2$ :  $z_2 \left[ \frac{2}{k_B} - \frac{2}{D} x_1 + \frac{1}{2D} (z_2 + 2x_1) \right] = \frac{4}{k_B} x_1$ . The only physical solution is indeed  $z_2 = 2x_1$ . Now assume that the induction hypothesis holds true for all  $i$  such that  $2 \leq i < j$ . Then for

$j$  odd:

$$\begin{aligned}
 z_j \left[ \frac{2}{k_B} - \frac{2}{D} \sum_{m=1}^{\frac{j-1}{2}} z_{2m} + \frac{1}{2D} (z_j + 2x_{j-1}) \right] &= \frac{4}{k_B} \left( x_{j-1} + 2 \sum_{i=1}^{j-2} (-1)^{j-1-i} x_i \right) \Rightarrow \\
 z_j \left[ \frac{2}{k_B} - \frac{2}{D} \frac{j-1}{2} 2x_1 + \frac{1}{2D} (z_j + 2(2(j-1) - 1)x_1) \right] &= \\
 \frac{4}{k_B} \left( (2(j-1) - 1)x_1 + 2 \sum_{i=1}^{j-2} (-1)^{j-1-i} (2i-1)x_1 \right) \Rightarrow \\
 z_j &= 2x_1
 \end{aligned}$$

For  $j$  even, the same procedure also results in  $z_j = 2x_1$ . This concludes the induction argument. Now we have for  $1 \leq j < n_p$ :  $x_j = (2j-1)x_1$ . Lastly we focus on  $j = n_p$ : first note that we can now simplify Equation 4.8 to:  $x_{n_p} = \frac{L}{2} + (n_p - 1)x_1$ . This means that  $z_{n_p} = x_{n_p} - x_{n_p-1} = \frac{L}{2} + (2 - n_p)x_1$ . Finally, by flux balance (Equation 4.5):  $k_B A(x_{n_p}) = J - k_B \sum_{i=1}^{n_p-1} A(x_i)$ . Invoking continuity (Equation 4.7) for  $A(x_{n_p})$  and replacing  $\sum_{i=1}^{n_p-1} A(x_i)$  for plasmid spacings, we obtain:

$$\begin{aligned}
 \frac{L}{k_B} &= \begin{cases} \left[ \left[ \frac{2}{k_B} + \frac{2}{D} z_{n_p} \right] \left[ \sum_{m=1}^{\frac{n_p}{2}-1} z_{2m+1} + x_1 \right] - \frac{1}{2D} z_{n_p} (z_{n_p} + 2x_{n_p-1}) + \frac{2}{k_B} x_1 \right. & \text{if } n_p \text{ even} \\ \left[ \left[ \frac{2}{k_B} + \frac{2}{D} z_{n_p} \right] \sum_{m=1}^{\frac{n_p-1}{2}} z_{2m} - \frac{1}{2D} z_{n_p} (z_{n_p} + 2x_{n_p-1}) + \frac{2}{k_B} x_1 \right. & \text{if } n_p \text{ odd} \end{cases} \\
 \Rightarrow \frac{L}{k_B} &= \left[ \frac{2}{k_B} + \frac{2}{D} z_{n_p} \right] (n_p - 1)x_1 - \frac{1}{2D} z_{n_p} (z_{n_p} + 2(2(n_p - 1) - 1)x_1) + \frac{2}{k_B} x_1
 \end{aligned}$$

Now we insert our expression  $z_{n_p} = \frac{L}{2} + (2 - n_p)x_1$  to find that  $x_1 = \frac{L}{2n_p}$ . We conclude that a symmetric ParA concentration leads to equal plasmid spacing:  $1 \leq j \leq n_p$ :  $x_j = \frac{L}{2n_p} + (j-1)\frac{L}{n_p}$ .

#### 4.4.2 Total ParA fluorescence analysis

Using the same data set as for the ParA asymmetry analysis, we also computed the total ParA-GFP intensity in every cell by summing over the linear projection values along the long axis. The cell volume (in pixels<sup>3</sup>) was determined by MicrobeTracker. We then multiplied this by the pixel size cubed to compute the total volume. Results are shown in Figure 4.18A.

#### 4.4.3 Total ParB fluorescence analysis

Using the same data set (consisting of single confocal planes) as for the determination of plasmid foci positioning in WT cells, we com-

puted the total ParB-GFP intensity per cell as described in the previous section. Results are shown in Figure 4.18B and are similar for summed ParB-GFP Z-stacks.

#### 4.4.4 *Diffusion/Immobilization model*

Further to the description of the diffusion/immobilization model in subsection 4.2.3 we describe here more details. The ParA concentration is assumed to be constant throughout the cell cycle, consistent with the total ParA-GFP fluorescence as a function of cell volume when expressed from an inducible promoter (Figure 4.18A). In accordance with estimates for average ParA copy numbers obtained by semi-quantitative Western blots [271], the ParA concentration is assumed to be 2400 ParA (dimers) per  $\mu\text{m}$  of nucleoid. Simulations start at time  $t = 0$  and run until time  $t$ , updated according to the Gillespie algorithm, exceeds a predefined time  $T$ . To simulate nucleoid growth during the cell cycle the nucleoid lattice is extended by two sites of size  $dx$  (not containing any ParA or plasmids), at one randomly chosen position along the nucleoid length. Such a growth event occurs at regular time intervals. Reaction propensities are then updated in accordance with the new state.

In Figure 4.3B the nucleoid grows from  $1.5\mu\text{m}$  to  $3\mu\text{m}$  in  $T = 40\text{min}$ , reflecting one cell cycle. Initially a quarter of the total ParA in the system is in the cytoplasmic ParA-ADP form, 11 ParA-ATP are bound to each plasmid to ensure initial anchoring, and the rest are bound randomly to the nucleoid. In Figure 4.3B the plasmid is initially located at site 0. In the simulations used to generate the histograms shown in Figure 4.3C, all plasmids are initially distributed randomly across the nucleoid. At regular time intervals of 5s the simulation state is output along with the plasmid positions to generate a time-averaged probability distribution for the plasmid positions along the long axis of the cell. In cases where the total number of plasmids ( $n_p$ ) is more than one, the plasmids are ordered and labeled  $1 \dots n_p$  according to their positions (by increasing site number) along the nucleoid. Their position is then used to generate distributions for every plasmid label  $1 \dots n_p$  for that particular overall number of plasmids  $n_p$ .

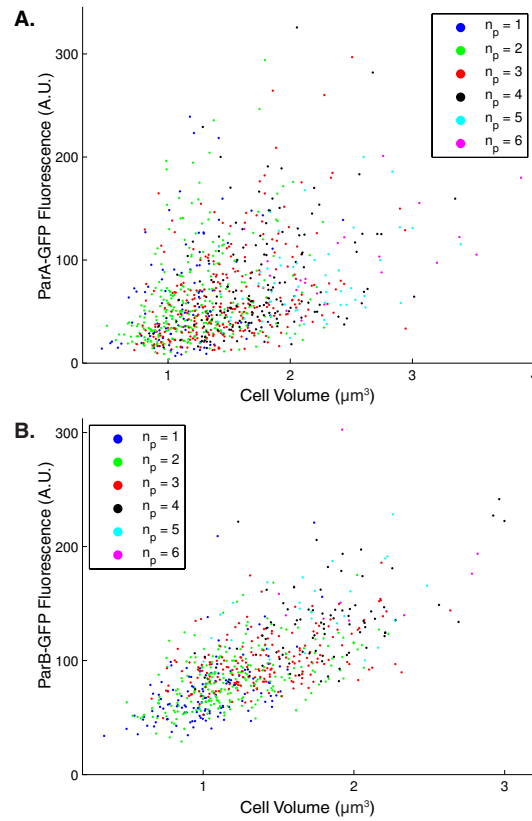


Figure 4.18: *par2* protein expression levels. (A) Scatter plot of ParA-GFP total fluorescence signal in single WT cells as a function of cell volume, when expressed from an inducible promoter ( $P_{lac}$ ). The different color labels indicate the number of plasmid foci. Plasmids: pSR233 (mini-R1, *par2*+,  $P_{lac}$ ::*parA*::eGFP, *tetO120*) and pSR124 ( $P_{BAD}$ ::*tetR*::mCherry). (B) Scatter plot of ParB-GFP total fluorescence signal in single WT cells as a function of cell volume, when expressed from its native promoter. Plasmid: pFS21 (*parC1*+, *parA*+, *parB*::sfGFP, *parC2*+); color labeling as in (A). Experiments performed by Florian Szardenings.

In the event of plasmid duplication at a particular site where an existing plasmid is located, a new plasmid without any bound ParA is added to the same site and the reaction propensities are updated accordingly. In case of two or more existing plasmids, one is chosen randomly for duplication. Plasmid duplication events in Figure 4.3B occur at regular time intervals  $T/3$ , although the model behaves equally well with duplication at any time as it dynamically segregates the plasmids to equally spaced positions.

#### 4.4.5 *Directed motion model*

The long axis could grow from  $1.5\mu m$  to  $3\mu m$  in length in  $T = 40min$  in length, while the short axis of the nucleoid lattice remained fixed. For wild-type directed motion model simulations the short axis length was  $30nm$  (directed motion model with short polymers) and  $25nm$  (directed motion model with long polymers). In the perturbed-nucleoid simulations, mobile ParA-ATP can diffuse past a plasmid with 10% (short) or 100% (long) of the normal diffusion rate and the short axis length of the nucleoid is altered to  $10nm$  in the long polymer model. Lastly, to allow for mobile ParA-ATP to move past the plasmid without being hydrolyzed,  $k_{mB}$  is reduced 10-fold compared to its standard value.

As for the diffusion/immobilization model, the total ParA concentration was constrained to be 2400 ParA (dimers) per  $\mu m$  of nucleoid (long axis) and the total length of simulated time was  $T = 40min$ . Initially a quarter of the total number of ParA in the system was in the cytoplasmic ParA-ADP form, with the rest distributed randomly on the nucleoid in the mobile ParA-ATP form. Initial plasmid positioning, state output, plasmid position distribution generation and plasmid duplication rules were also as described above. Nucleoid growth was implemented as described previously, with one generalization: a position along the long axis of the nucleoid was first chosen randomly. Then two nucleoid slices of 1 site (along the long axis) by  $S$  sites (along the short axis) were inserted.

#### 4.4.6 *Three dimensional nucleoid and ParA structure analysis*

To compare the extent of overlay and 3D structure of Hoechst (nucleoid DNA) stain and ParA-GFP, we first had to align the appropriate Z-stack pairs in an unbiased manner. To achieve this, one phase contrast (PC) image (at mid  $z$  height) of the Hoechst signal sections was aligned with one GFP section PC image (at the same  $z$  position) using the TurboReg ImageJ plugin (option: translation) [279], after cropping both PC images to match the output size of the deconvolved Z-stacks. Using the same translation as for the Hoechst PC image, Hoechst Z-stacks were then translated in ImageJ to align them with the ParA-GFP Z-stacks. We determined cell outlines in MT as described above using the PC image acquired with the GFP channel and excluded cells that did not show visible ParA-GFP and Hoechst stain simultaneously. We then computed the linear distributions (for every  $z$  height) along the long cell axis for the deconvolved Hoechst and ParA-GFP Z-stacks. We next determined for the ParA-GFP and Hoechst signals separately in every cell the maximal intensity value in the whole cell ( $I_{maxcell}$ ) and the maximal values at every  $z$  height ( $I_{max}(z)$ ). To find the 9  $z$  planes from the Z-stacks ( $dz = 0.1\mu m$ ) that are in-focus for each cell in an automated fashion, we summed  $I_{max}(z)$  over 9 consecutive  $z$  positions including a given starting plane and determined the starting plane that gave the largest associated summed value. This starting plane and its 8 consecutive planes formed the in focus plane set. We verified that this method generated the right focus planes by inspecting the chosen planes visually for several cells. This method circumvents the problem of different focus planes for cells on the same image stack as well as alignment inaccuracies in the  $z$  direction between ParA-GFP and Hoechst signals which are difficult to control for manually.

To determine the fraction of ParA-GFP intensity signal that overlaps with Hoechst signal and vice versa we computed Manders overlap coefficients [278]. This method requires a choice of threshold  $T_{Manders}$  to distinguish between positions  $(x, z)$  that are considered to contain or lack sufficient intensity signal. We therefore performed our analyses for the complete range of threshold values to show that our



qualitative conclusions are insensitive to the choice of a particular  $T_{Manders}$  (Figure 4.10C, Figure 4.12D). Manders overlap coefficients of ParA-GFP and Hoechst were calculated as follows:

$$M_{ParA-GFP} = \frac{\sum_{x,z} I_{ParA-GFP,coloc}(x, z)}{\sum_{x,z} I_{ParA-GFP}(x, z)}$$

with

$$I_{ParA-GFP,coloc}(x, z) = \begin{cases} I_{ParA-GFP}(x, z) & \text{if } \frac{I_{Hoechst}(x,z)}{I_{Hoechst,maxcell}} \geq T_{Manders} \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

Likewise the Manders overlap coefficient of Hoechst onto ParA-GFP is defined as:

$$M_{Hoechst} = \frac{\sum_{x,z} I_{Hoechst,coloc}(x, z)}{\sum_{x,z} I_{Hoechst}(x, z)}$$

with

$$I_{Hoechst,coloc}(x, z) = \begin{cases} I_{Hoechst}(x, z) & \text{if } \frac{I_{ParA-GFP}(x,z)}{I_{ParA-GFP,maxcell}} \geq T_{Manders} \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

Note that taking  $T_{Manders} = 0$ , will generate an overlap coefficient of one by construction. The normalization to  $I_{maxcell}$  in determining the colocalizing positions allows the overlap coefficients to be comparable between cells.

In a small fraction of cells the alignment procedure described above did not result in proper alignment. This is clearly reflected in the correlation coefficient  $r$  values being considerably lower for these cases than for the cell population mean  $r$  value. However, without excluding these few, possibly false negative, cases the population mean  $r$  value is still high (0.81 and 0.68 for WT and Nal-treated cells respectively), indicating that ParA-GFP and Hoechst signals generally correlate strongly at a population level. Poor alignment affects Manders overlap coefficients for the ParA-GFP and Hoechst signals on average equally and is not biased towards a particular strain/treatment. Therefore the observed misalignment of a small fraction of cells does not affect the qualitative conclusions that we state in this study.

Parameter	Description	Value	Notes
$D_A$	Nucleoid bound ParA-ATP diffusion constant	$1\mu m^2/s$	Constrained by experiment [180],  value can be varied by several orders of magnitude without loss of qualitative behaviour of system. Note that this form of ParA-ATP does not have an effect on the mobility of plasmids, since only polymeric ParA-ATP, immobile due to the interaction with the nucleoid, can direct the motion of a plasmid.
$D_P$	Plasmid diffusion constant	$3 \cdot 10^{-4}\mu m^2/s$	Constrained by experiment (Figure 4.4C).
$k_{on}$	ParA-ATP nucleoid binding	$50s^{-1}$	Constrained by experiment [166].
$k_B$	Plasmid bound polymeric ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$68.5s^{-1}$	Chosen to be the same as $k_B$ in diffusion/immobilization model for consistency; constrained by $k_B \gg k_{dm}$ which ensures that all the ParA-ATP at the location of a plasmid is converted into cytoplasmic ParA-ADP before the plasmid moves to a neighboring site. Value can be varied within a wide range without loss of qualitative behaviour of system.
$k_{mB}$	Plasmid bound mobile ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$40s^{-1}$	Fitted, value can be varied within a wide range without loss of qualitative behaviour of system. Setting this rate too high depletes ParA-ATP locally around a plasmid, which inhibits directed plasmid motion events.
$k_W$	(Cytoplasmic) ParA-ADP to ParA-ATP conversion	$1/15s^{-1}$	Constrained by experiment [166], this value should be low enough to ensure that cytoplasmic ParA diffusion can generate a uniform cytoplasmic ParA-ATP and ParA-ADP concentration.
$k_{dm}$	Plasmid directed motion rate (in presence of one neighboring plasmid)	$0.8s^{-1}$	Constrained by experiment (Figure 4.4A). If interpreted as biased plasmid diffusion along the polymer, this would result effectively in a maximal plasmid diffusion constant of $1 \cdot 10^{-4}\mu m^2/s$ (short) and $1.2 \cdot 10^{-4}\mu m^2/s$ (long). These values are consistent with the free diffusion constant $D_P$ (see above), since the interaction with immobile ParA-ATP polymers could lower the plasmid mobility.
$k_p$	Polymerization: mobile ParA-ATP to polymeric ParA-ATP conversion	$800s^{-1}$ (short) $10^6s^{-1}$ (long)	Fitted together with $k_{dm}$ and $k_W$ to ensure equal plasmid spacing.  $k_p$ and $k_{dp}$ together with the total ParA-ATP concentration determine the extent of ParA-ATP polymerization.
$k_{dp}$	Spontaneous depolymerization: ParA-ATP to mobile ParA-ATP conversion	$10s^{-1}$ (short) $10^{-4}s^{-1}$ (long)	Fitted. See notes on $k_p$ parameter above.
$S$	Short axis length of the nucleoid region where nucleoid bound ParA-ATP can polymerize.	$30nm$ (short) $25nm$ (long)	Fitted, values should be small compared to the long nucleoid axis length to ensure that segregation occurs along the long nucleoid axis.
	Perturbed nucleoid simulations		Parameter values as above unless specified below. See also subsection 4.2.5 for further details.
$k_{mB}$	Plasmid bound mobile ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$4s^{-1}$	Fitted, value is chosen to simulate the effect of a disordered nucleoid structure, allowing mobile ParA-ATP to diffuse past plasmids.
$S$	Short axis length of the nucleoid region where nucleoid bound ParA-ATP can polymerize.	$30nm$ (short) $10nm$ (long)	Fitted, values are chosen to ensure a sufficient amount of mobile ParA-ATP.

Table 4.4: Parameter values used in the directed motion models, short (long) indicates short (long) polymer model.

## CONCLUSIONS & DISCUSSION

---

# 5

To conclude this thesis we summarize how through cycles of mathematical modelling and experiments we have advanced our quantitative understanding of two separate topics: transcriptional regulation of the floral repressor *FLOWERING LOCUS C* (*FLC*) in *Arabidopsis thaliana* and spatial positioning of low copy number plasmids in *Escherichia coli*. Despite the diversity in biological subjects, the modelling approach in the two different projects showed many similarities. Therefore I conclude this thesis with a discussion of the commonalities in theoretical methodology used in my research.

### 5.1 Autonomous pathway regulation of *FLC*

A major aim of this PhD project has been to quantitatively dissect *FLC* regulation by the autonomous pathway using mathematical modelling. A qualitative mechanism in the form of chromatin silencing by FLD and FCA promoting *COOLAIR* proximal polyadenylation formed the starting point of my research [98, 114, 120]. The kinetic coupling between splicing and RNA Polymerase II (Pol II) elongation rates was suggested as a hypothesis to explain the differential upregulation of the proximal and distal *COOLAIR* isoforms [98, 127]. However, it was unclear what caused the sense *FLC* expression upregulation in autonomous pathway mutants. Several initial analytical models, based on regulation of transcription initiation, premature termination or nascent RNA degradation were able to explain the existing *FLC* fold upregulations. Pol II Chromatin immunoprecipitation (ChIP) experiments then qualitatively falsified models based on premature termination. Surprisingly, Pol II levels showed little increase across the gene, which led to the hypothesis of coordinated initiation and elongation. This also naturally explained the differential *COOLAIR* isoform upregulation in *fca-9* and *fld-4* mutants. Both a relatively simple analytical model and more detailed mechanistic stochastic simulations could reproduce this Pol II upregulation

alongside the observed RNA upregulations. A further mathematical analysis successfully predicted the chromatin-bound and total sense intron1 RNA upregulation, thus providing an experimental validation for the coordination model. In order to fit the available data, the model predicts similar coordinated changes to occur for antisense transcription, however we have not been able to validate these predictions specifically for the antisense strand.

Nascent 5'-3' RNA degradation had initially been an alternative to the coordination model, however in addition to its qualitative invalidation by Pol II ChIP, it could not quantitatively fit the data either, unlike the coordination model. An expression analysis of *xrn* mutants provided further genetic evidence against this model. 3'-5' degradation after endonucleolytic cleavage of Pol II-bound RNA has also been investigated. However besides the Pol II upregulation being inconsistent, an extensive experimental mapping of RNA 3' end sided against this mechanism playing a major role in *FLC* regulation. RNA degradation at the locus after elongation is also quantitatively inconsistent with the experimental data for the same reasons as the nascent 5'-3' RNA degradation model. Our chromatin-bound RNA data reconfirms the assumption that *FLC* upregulation is chromatin-based, ruling out any post-transcriptional mechanisms. We conclude that we have found decisive experimental evidence in favour of a model with coordinated initiation and elongation rate changes and systematically ruled out alternative, *a priori* possible modes of regulation.

Despite continued efforts we have not provided sufficient evidence to elucidate the molecular mechanism behind coordination of transcriptional initiation and elongation. Mechanistic simulations based on FCA-mediated *COOLAIR* proximal termination leading to FLD activity at the *FLC* locus are quantitatively consistent with all available data. However, the exact connection between FCA and FLD activity remains unclear. In the author's opinion, these questions are, at least in the short term, better addressed through further biochemical and genetic investigations. It is possible that the R-loop, found around the proximal *COOLAIR* polyadenylation site, is involved in

this process. Also, how FLD activity affects transcriptional coordination remains to be resolved. It likely works through histone demethylation leading to an altered chromatin state that then affects initiation and elongation together. However, we cannot currently rule out if FLD has other targets that affect *FLC* regulation. *COOLAIR* splicing and transcription are clearly functionally important for repression [127, 128]. However, *FLC* upregulation in an *fca-9* mutant is not epistatic to *COOLAIR* disruption (Z. Wu & C. Dean, unpublished). So the involvement of *COOLAIR* in autonomous pathway repression remains to be fully understood. We conclude that this thesis provides evidence supportive of a role for chromatin and *COOLAIR* in *FLC* repression by the autonomous pathway.

## 5.2 Cellular variation of *FLC* expression

Our investigations of cellular *FLC* variability were initially inspired by a mechanism found in yeast whereby non-coding RNA transcription toggles, resulting in variegated gene expression [63, 145]. At *FLC* the proximal and distal poly(A) sites could then provide such a toggle, potentially with opposite effects on sense expression: the proximal site would enhance *FLC* repression while the distal site would enhance *FLC* activation [127, 256]. Indeed we found that *FLC* shows a broad range of cellular expression levels. However, we also found that cellular volume correlates strongly with RNA levels. The magnitude of the residual variance indicates that, for a given volume, transcription is well described by a Poisson process. This contrasts with variegated expression as found for the genetic toggle mechanism in yeast. Instead it suggests that the cellular *FLC* concentration is tightly maintained at a constant value. This could be important to prevent a scenario where a low cellular *FLC* concentration due to fluctuations would lead to an untimely induction of floral activators such as *FT*, whose protein is mobile between cells [231–233]. Such premature activation of floral activators in certain parts of the plant, for instance the leaves, could then have consequences on the timing of the floral transition in the shoot apical meristem.

Our results also shed light on the connection between autonomous pathway repression and Polycomb repression, which is critical for

epigenetic silencing after prolonged cold. It was *a priori* possible that cell autonomous epigenetic silencing by Polycomb would also occur in non-vernalized (NV) conditions. This should then result in a bimodal *FLC* concentration distribution. Genetic evidence indicates that Polycomb does contribute to NV *FLC* repression [280], but without affecting the functionality of the autonomous pathway (H. Yang & C. Dean, unpublished). To reconcile these features, it could be that prior to prolonged cold, the positive feedback required for epigenetic silencing by Polycomb is absent. This would then result in an analog NV repression by both the autonomous pathway and Polycomb, whilst the cold induces the required feedback to ensure maintainance of epigenetic silencing by Polycomb after cold. A combination of further experimental and theoretical investigations could further elucidate how the actions of different pathways are integrated at the locus before, during and after cold to generate a comprehensive understanding of *FLC* regulation. This synthesis is important for an overall understanding of flowering time in *Arabidopsis*. Flowering is of course an important trait for crop breeders. The development of crop varieties, either through conventional breeding technologies or genetic modification, with a quantitatively tuned flowering time through modulation of conserved *FLC* regulators such as FRIGIDA and the autonomous pathway, could help to further improve crop yields.

### 5.3 Equal plasmid spacing over the nucleoid

Regular plasmid spacing has been known to occur for several low copy number plasmids in a variety of prokaryotes for a over decade [155]. Despite substantial genetic and biochemical investigations, it remained unclear how such equal spacing was generated [159]. At the time this project started, in the autumn of 2010 for my Masters research project [153], several competing qualitative mechanisms were posed in the literature [159, 166, 173, 184]: some based on ParA polymerization, others on a gradient of ParA. Furthermore several *parABC* loci exhibit oscillations [159, 175], whilst others do not [166, 173]. It was also unclear whether polymerization was important for segregation [184]. During my masters project we developed initial versions of the diffusion/immobilization model and the long

polymer model (with directed motion) and showed that both could position plasmids under certain conditions [153]. The experiments and further modelling, including all final model versions that were eventually published in [271] and described in this thesis, were performed during my PhD research. We showed that the essential components responsible for equal plasmid spacing are non-specific ParA-ATP nucleoid binding, ParA-ATP diffusion over the nucleoid and ParA-ATP to ADP conversion predominantly locally at the plasmid through the ParB-*parC* complex. We realized through the analytical mathematical analysis that the means by which plasmids move is not critical to achieve equal spacing, as long as plasmids can displace toward a higher ParA concentration.

The strongest prediction of the diffusion/immobilization model was that it required a high intrinsic mobility of plasmids. Presence of an active *parABC* segregation system would then only slow down plasmid movement. By a time series monitoring of plasmid positions over time, we found instead that the *parABC* system activates plasmid mobility. This falsified the diffusion/immobilization and favoured the polymer model. Critical referees rightfully commented that we assumed ParA polymers to be responsible for pulling plasmids, in line with earlier work [159]. However the minimal requirement was only that ParA can direct plasmid motion on a small length scale. Polymers pulling was only one specific mechanism that could generate this directed motion. So the degree of polymerization is not essential for positioning and we relabeled our polymer model to the directed motion model, without further specifying what underlying mechanism would cause the directed motion. We simulated two versions of directed motion, one with short polymers and one with strong polymerization kinetics, to clarify that our model is general enough to explain positioning in several segregation systems with varying degrees of polymerization.

Our modelling had also suggested that ParA competition, rather than oscillations were key to positioning. This was quite a surprise at the time since pB171 ParA had been repeatedly shown to oscillate [156, 159]. A systematic unbiased experimental analysis confirmed, how-

ever, that the ParA asymmetry was relatively low, consistent with the ParA competition mechanism and contrasting with continued oscillations. We resolved this paradox by showing that oscillations do occur when plasmids are not stably positioned. Rather plasmids move along in the wake of retracting ParA structures, a behaviour that can give rise to pole-to-pole ParA oscillations. However once plasmids are stably positioned, ParA asymmetry is low and oscillations cease.

To maintain a concentration gradient over the nucleoid that can specify equal plasmid spacing, the one dimensionality of ParA structures seems important. It remains unclear what constitutes these ParA structures, at least for plasmid pB171: long polymer bundles, a multitude of shorter filaments, or no polymerization at all *in vivo*. The latter would contrast with *in vitro* polymerization of pB171 ParA [158]. We also showed that plasmid positioning depends on the nucleoid structure, likely through an effect on ParA structure formation. It remains an interesting and open question how nucleoid organization mechanistically connects to ParA and plasmid dynamics.

## 5.4 Mathematical modelling methodology

In both the plant and bacterial projects, the starting point of my modelling has been an experimental observation that required a quantitative explanation. For the *FLC* project this was the RNA fold up-regulation in autonomous pathway mutants, whilst for the plasmids it was the key observation that plasmids are spaced equally along the long cell axis irrespective of cell length or plasmid copy number [159]. Throughout both projects I made use of similar spatiotemporal modelling techniques such as partial differential equations, the theory of stochastic processes and numerical Gillespie simulations to perform detailed mechanistic modelling. In both cases we repeatedly utilized the considerable prior knowledge base of genetic and biochemical evidence that had already elucidated a qualitative understanding of several aspects of the mechanisms. Based on that information, we developed dynamic models, that were then connected and compared with the often non-dynamic experimental evidence by assuming steady state conditions. This resulted in multiple ini-



tial models that could explain the available data. For the plasmids we had the diffusion/immobilization model competing with directed motion to explain the mode of plasmid movement. Also, *FLC* expression could be regulated through transcriptional initiation, premature termination or nascent RNA degradation. Subsequent experiments were needed to distinguish between the models: the plasmid mobility assay and Pol II ChIP. Models were then adjusted or refined where needed to be made consistent with the new observations and then utilized to make further key model predictions. These predictions were then experimentally validated where possible. The ParA asymmetry measurements validated that competing ParA structures, rather than oscillations were key to plasmid positioning. The intronic RNA experiments confirmed the elongation rate changes, a novel feature of transcriptional regulation in plants. We conclude that iterative cycles of modelling and experiments have substantially advanced our quantitative understanding of plasmid spacing over the nucleoid and *FLC* transcriptional regulation.

## 5.5 Outlook

Over the last decade predictive mathematical modelling in biology has matured as a scientific discipline. It is increasingly accepted to use mathematical analysis as part of a study. However, much progress is still to be made to make biology a truly quantitative science. As stated above, predictive modelling seems most fruitful when it goes hand in hand with experiments. In the author's opinion this is where both its strengths and weaknesses lie. Strengths because mathematics can be used to generate counterintuitive predictions that in combination with experimental validation can lead to novel findings. However, scientific advancement proceeds through convincing fellow scientists that the presented evidence is in favour of a particular hypothesis. To convince a less mathematically sophisticated scientific peer of a mathematical argument, or even the value of certain experimental data as evidence in favour of a mathematical model, forms a challenge. Presence of a larger community of quantitative biologists should help to reduce this problem. Moreover, further mathematics education is needed to better equip future generations of biologists. Likewise for scientists with a quantitative background like myself,

sufficient domain knowledge and interest in biology is essential in order to make a useful contribution to the field.

Skeptics may ask what benefits a quantitative understanding brings over a description in words. I would like to counter this argument with the question: in what case is a qualitative understanding sufficient? Not to build a nuclear power plant, a train, a house, a car, a computer, or even a telephone. For nearly all present day technology, a mathematical description is essential. I do not see why biology and its applications, such as food production and disease treatment, should be any different. Some say the complexity in biology is such that quantitative understanding is still beyond the reach of our ability. I am convinced that it only takes interest and a quantitative mindset to perform the research and technological development required for progress in this area. I hope to have convinced the reader that the results described in this thesis form an example of a small step towards a quantitative understanding of life.

## BIBLIOGRAPHY

---

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter. **Molecular biology of the cell**. 5th ed. Garland Science, 2008 (see pp. 15–17, 19–22, 37, 38, 44).
- [2] A. J. Herr, M. B. Jensen, T. Dalmay and D. C. Baulcombe. **RNA polymerase IV directs silencing of endogenous DNA**. *Science* 308.5718 (2005), 118–120 (see pp. 15, 16).
- [3] T. S. Ream, J. R. Haag, A. T. Wierzbicki, C. D. Nicora, A. D. Norbeck, J. K. Zhu, G. Hagen, T. J. Guilfoyle, L. Pasa-Tolic and C. S. Pikaard. **Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA Polymerase II**. *Molecular Cell* 33.2 (2009), 192–203 (see p. 16).
- [4] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker and E. Heard. **Spatial partitioning of the regulatory landscape of the X-inactivation centre**. *Nature* 485.7398 (2012), 381–385 (see p. 16).
- [5] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren. **Topological domains in mammalian genomes identified by analysis of chromatin interactions**. *Nature* 485.7398 (2012), 376–380 (see p. 16).
- [6] E. M. Blackwood and J. T. Kadonaga. **Going the distance: a current view of enhancer action**. *Science* 281.5373 (1998), 60–63 (see p. 16).
- [7] T. I. Lee and R. A. Young. **Transcriptional regulation and its misregulation in disease**. *Cell* 152.6 (2013), 1237–1251 (see pp. 16, 17, 19).
- [8] S. Henikoff and A. Shilatifard. **Histone modification: cause or cog?** *Trends in Genetics* 27.10 (2011), 389–396 (see p. 16).
- [9] L. A. Selth, S. Sigurdsson and J. Q. Svejstrup. **Transcript elongation by RNA Polymerase II**. *Annual Review of Biochemistry* 79 (2010), 271–293 (see pp. 17, 19).
- [10] J. Q. Svejstrup. **The RNA Polymerase II transcription cycle: cycling through chromatin**. *Biochimica Et Biophysica Acta* 1677.1-3 (2004), 64–73 (see p. 17).
- [11] A. Johnson, R. H. Wu, M. Peetz, S. P. Gygi and D. Moazed. **Heterochromatic gene silencing by activator interference and a transcription elongation barrier**. *Journal of Biological Chemistry* 288.40 (2013), 28771–28782 (see p. 17).
- [12] S. J. Petesch and J. T. Lis. **Overcoming the nucleosome barrier during transcript elongation**. *Trends in Genetics* 28.6 (2012), 285–294 (see p. 17).
- [13] M. L. Phelan, S. Sif, G. J. Narlikar and R. E. Kingston. **Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits**. *Molecular Cell* 3.2 (1999), 247–253 (see p. 17).
- [14] T. I. Lee and R. A. Young. **Transcription of eukaryotic protein-coding genes**. *Annual Review of Genetics* 34 (2000), 77–137 (see p. 17).

- [15] J. P. Hsin and J. L. Manley. **The RNA Polymerase II CTD coordinates transcription and RNA processing.** *Genes & Development* 26.19 (2012), 2119–2137 (see p. 17).
- [16] S. Buratowski. **Progression through the RNA Polymerase II CTD cycle.** *Molecular Cell* 36.4 (2009), 541–546 (see pp. 17, 72).
- [17] A. Mayer, M. Lidschreiber, M. Siebert, K. Leike, J. Soding and P. Cramer. **Uniform transitions of the general RNA Polymerase II transcription complex.** *Nature Structural & Molecular Biology* 17.10 (2010), 1272–+ (see pp. 17, 19, 93).
- [18] P. Komarnitsky, E. J. Cho and S. Buratowski. **Different phosphorylated forms of RNA Polymerase II and associated mRNA processing factors during transcription.** *Genes & Development* 14.19 (2000), 2452–2460 (see p. 17).
- [19] B. Li, M. Carey and J. L. Workman. **The role of chromatin during transcription.** *Cell* 128.4 (2007), 707–719 (see pp. 18, 19).
- [20] H. H. Ng, F. Robert, R. A. Young and K. Struhl. **Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.** *Molecular Cell* 11.3 (2003), 709–719 (see pp. 17, 135).
- [21] E. Bernstein and C. D. Allis. **RNA meets chromatin.** *Genes & Development* (2005), 1635–1655 (see p. 17).
- [22] J. Dover, J. Schneider, M. A. Tawiah-Boateng, A. Wood, K. Dean, M. Johnston and A. Shilatifard. **Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.** *Journal of Biological Chemistry* 277.32 (2002), 28368–28371 (see p. 17).
- [23] Z. W. Sun and C. D. Allis. **Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast.** *Nature* 418.6893 (2002), 104–108 (see p. 17).
- [24] A. J. Ruthenburg, C. D. Allis and J. Wysocka. **Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark.** *Molecular Cell* 25.1 (2007), 15–30 (see p. 17).
- [25] R. J. Sims, S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage, P. Tempst, J. L. Manley and D. Reinberg. **Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing.** *Molecular Cell* 28.4 (2007), 665–676 (see p. 17).
- [26] X. Y. Zhang, Y. V. Bernatavichute, S. Cokus, M. Pellegrini and S. E. Jacobsen. **Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*.** *Genome Biology* 10.6 (2009) (see p. 17).
- [27] T. Margaritis, V. Oreal, N. Brabers, L. Maestroni, A. Vitaliano-Prunier, J. J. Benschop, S. van Hooff, D. van Leenen, C. Dargemont, V. Geli and F. C. P. Holstege. **Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription.** *PLoS Genetics* 8.9 (2012) (see p. 17).

- [28] T. Kim, Z. Y. X. Sandra, S. Clauder-Munster, L. M. Steinmetz and S. Buratowski. **Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics.** *Cell* 150.6 (2012), 1158–1169 (see pp. 17, 19).
- [29] T. Kim and S. Buratowski. **Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions.** *Cell* 137.2 (2009), 259–272 (see pp. 17, 24, 29).
- [30] M. J. Carrozza, B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn and J. L. Workman. **Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3s to suppress spurious intragenic transcription.** *Cell* 123.4 (2005), 581–592 (see pp. 19, 95).
- [31] M. C. Keogh, S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Y. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski and N. J. Krogan. **Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex.** *Cell* 123.4 (2005), 593–605 (see pp. 19, 95).
- [32] M. Smolle, S. Venkatesh, M. M. Gogol, H. Li, Y. Zhang, L. Florens, M. P. Washburn and J. L. Workman. **Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange.** *Nature Structural & Molecular Biology* 19.9 (2012), 884–892 (see p. 19).
- [33] P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X. S. Liu and J. Ahringer. **Differential chromatin marking of introns and expressed exons by H3K36me3.** *Nature Genetics* 41.3 (2009), 376–381 (see p. 19).
- [34] A. Rechtsteiner, S. Ercan, T. Takasaki, T. M. Phippen, T. A. Egelhofer, W. C. Wang, H. Kimura, J. D. Lieb and S. Strome. **The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny.** *PLoS Genetics* 6.9 (2010) (see p. 19).
- [35] H. F. Qiu, C. H. Hu and A. G. Hinnebusch. **Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters.** *Molecular Cell* 33.6 (2009), 752–762 (see p. 19).
- [36] L. S. Churchman and J. S. Weissman. **Nascent transcript sequencing visualizes transcription at nucleotide resolution.** *Nature* 469.7330 (2011), 368–373 (see pp. 19, 52, 93).
- [37] S. Nechaev, D. C. Fargo, G. dos Santos, L. W. Liu, Y. Gao and K. Adelman. **Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in drosophila.** *Science* 327.5963 (2010), 335–338 (see p. 19).
- [38] T. Henriques, D. A. Gilchrist, S. Nechaev, M. Bern, G. W. Muse, A. Burkholder, D. C. Fargo and K. Adelman. **Stable pausing by RNA Polymerase II provides an opportunity to target and integrate regulatory signals.** *Molecular Cell* 52.4 (2013), 517–528 (see pp. 19, 52).

- [39] M. S. Buckley, H. Kwak, W. R. Zipfel and J. T. Lis. **Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation.** *Genes & Development* 28.1 (2014), 14–19 (see pp. 19, 52, 71).
- [40] A. Mayer, J. di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos and L. S. Churchman. **Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution.** *Cell* 161.3 (2015) (see pp. 19, 20).
- [41] Y. Brody, N. Neufeld, N. Bieberstein, S. Z. Causse, E. M. Bohnlein, K. M. Neugebauer, X. Darzacq and Y. Shav-Tal. **The *in vivo* kinetics of RNA Polymerase II elongation during co-transcriptional splicing.** *PLoS Biology* 9.1 (2011) (see pp. 20, 21, 59, 61, 62, 65, 89, 103, 119, 125, 130).
- [42] S. F. de Almeida, A. R. Grosso, F. Koch, R. Fenouil, S. Carvalho, J. Andrade, H. Levezinho, M. Gut, D. Eick, I. Gut, J.-C. Andrau, P. Ferrier and M. Carmo-Fonseca. **Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36.** *Nature Structural & Molecular Biology* 18.9 (2011), 977–983 (see p. 20).
- [43] M. de la Mata, C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley and A. R. Kornblihtt. **A slow RNA Polymerase II affects alternative splicing *in vivo*.** *Molecular Cell* 12.2 (2003), 525–532 (see p. 20).
- [44] D. L. Bentley. **Coupling mRNA processing with transcription in time and space.** *Nature Reviews Genetics* 15.3 (2014), 163–175 (see pp. 20, 62, 64, 65, 69, 83, 85, 90, 103, 130).
- [45] C. Alen, N. A. Kent, H. S. Jones, J. O’Sullivan, A. Aranda and N. J. Proudfoot. **A role for chromatin remodeling in transcriptional termination by RNA Polymerase II.** *Molecular Cell* 10.6 (2002), 1441–1452 (see p. 20).
- [46] C. Dean, S. Tamaki, P. Dunsmuir, M. Favreau, C. Katayama, H. Dooner and J. Bedbrook. **Messenger-RNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*.** *Nucleic Acids Research* 14.5 (1986), 2229–2240 (see p. 20).
- [47] N. J. Proudfoot. **Ending the message: poly(A) signals then and now.** *Genes & Development* 25.17 (2011), 1770–1782 (see pp. 20, 52, 53, 74, 75).
- [48] M. Kim, N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedeá, J. F. Greenblatt and S. Buratowski. **The yeast Rat1 exonuclease promotes transcription termination by RNA Polymerase II.** *Nature* 432.7016 (2004), 517–22 (see p. 20).
- [49] S. West, N. Gromak and N. J. Proudfoot. **Human 5’ → 3’ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites.** *Nature* 432.7016 (2004), 522–5 (see p. 20).
- [50] S. West and N. J. Proudfoot. **Transcriptional termination enhances protein expression in human cells.** *Molecular Cell* 33.3 (2009), 354–364 (see p. 20).

- [51] T. Nojima, M. Dienstbier, S. Murphy, N. J. Proudfoot and M. J. Dye. **Definition of RNA Polymerase II CoTC terminator elements in the human genome.** *Cell Reports* 3.4 (2013), 1080–1092 (see p. 21).
- [52] F. Q. Liu, S. Bakht and C. Dean. **Cotranscriptional role for *Arabidopsis* DICER-LIKE 4 in transcription termination.** *Science* 335.6076 (2012), 1621–1623 (see p. 21).
- [53] D. Z. Hazelbaker, S. Marquardt, W. Wlotzka and S. Buratowski. **Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination.** *Molecular Cell* 49.1 (2013), 55–66 (see pp. 21, 69, 71).
- [54] A. Wagschal, E. Rousset, P. Basavarajaiah, X. Contreras, A. Harwig, S. Laurent-Chabalier, M. Nakamura, X. Chen, K. Zhang, O. Meziane, F. Boyer, H. Parrinello, B. Berkhout, C. Terzian, M. Benkirane and R. Kiernan. **Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII.** *Cell* 150.6 (2012), 1147–1157 (see p. 21).
- [55] F. Ozsolak, P. Kapranov, S. Foissac, S. W. Kim, E. Fishilevich, A. P. Monaghan, B. John and P. M. Milos. **Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation.** *Cell* 143.6 (2010), 1018–1029 (see p. 21).
- [56] D. Schulz, B. Schwalb, A. Kiesel, C. Baejen, P. Torkler, J. Gagneur, J. Soeding and P. Cramer. **Transcriptome surveillance by selective termination of noncoding RNA synthesis.** *Cell* 155.5 (2013), 1075–1087 (see p. 21).
- [57] K. Brannan, H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, K. Hansen, R. Davis, J. Lykke-Andersen and D. L. Bentley. **mRNA decapping factors and the exonuclease xrn2 function in widespread premature termination of RNA Polymerase II transcription.** *Molecular Cell* 46.3 (2012), 311–324 (see p. 21).
- [58] A. E. Almada, X. B. Wu, A. J. Kriz, C. B. Burge and P. A. Sharp. **Promoter directionality is controlled by U1 RNAPII and polyadenylation signals.** *Nature* 499.7458 (2013), 360–U141 (see pp. 21, 70).
- [59] S. Marquardt, R. Escalante-Chong, N. Pho, J. Wang, L. S. Churchman, M. Springer and S. Buratowski. **A chromatin-based mechanism for limiting divergent noncoding transcription.** *Cell* 157.7 (2014), 1712–1723 (see p. 21).
- [60] S. Masuda, R. Das, H. Cheng, E. Hurt, N. Dorman and R. Reed. **Recruitment of the human TREX complex to mRNA during splicing.** *Genes & Development* 19.13 (2005), 1512–1517 (see p. 21).
- [61] L. Davidson, A. Kerr and S. West. **Co-transcriptional degradation of aberrant pre-mRNA by xrn2.** *Embo Journal* 31.11 (2012), 2566–2578 (see pp. 21, 74).
- [62] J. Houseley and D. Tollervey. **The many pathways of RNA degradation.** *Cell* 136.4 (2009), 763–776 (see p. 21).
- [63] A. C. Tuck and D. Tollervey. **An RNA reset button.** *Molecular Cell* 45.4 (2012), 435–436 (see pp. 21, 33, 34, 201).

- [64] M. Schmid, M. B. Poulsen, P. Olszewski, V. Pelechano, C. Saguez, I. Gupta, L. M. Steinmetz, C. Moore and T. H. Jensen. **Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins.** *Molecular Cell* 47.2 (2012), 267–280 (see p. 21).
- [65] K. Yamada et al. **Empirical analysis of transcriptional activity in the *Arabidopsis* genome.** *Science* 302.5646 (2003), 842–846 (see p. 21).
- [66] P. Kapranov, A. T. Willingham and T. R. Gingeras. **Genome-wide transcription and the implications for genomic organization.** *Nature Reviews Genetics* 8.6 (2007), 413–423 (see p. 22).
- [67] J. E. Wilusz, H. Sunwoo and D. L. Spector. **Long noncoding RNAs : functional surprises from the RNA world** long noncoding RNAs : functional surprises from the RNA world. *Genes & Development* (2009), 1494–1504 (see p. 22).
- [68] C. P. Ponting, P. L. Oliver and W. Reik. **Evolution and functions of long noncoding RNAs.** *Cell* 136.4 (2009), 629–641 (see p. 22).
- [69] J. Wang, F. Lu, Q. Ren, H. Sun, Z. S. Xu, R. F. Lan, Y. Q. Liu, D. Ward, J. M. Quan, T. Ye and H. Zhang. **Novel histone demethylase lsd1 inhibitors selectively target cancer cells with pluripotent stem cell properties.** *Cancer Research* 71.23 (2011), 7238–7249 (see p. 22).
- [70] A. K. Eggleston, A. Eccleston, B. Marte and C. Lupp. **Regulatory RNA.** *Nature* 482.7385 (2012), 321 (see p. 22).
- [71] J. S. Mattick. **Rna regulation: a new genetics?** *Nature Reviews Genetics* 5.4 (2004), 316–323 (see p. 22).
- [72] M. Okamoto and M. Seki. **Expression profile and 5'-terminal structure of *Arabidopsis* antisense transcripts expressed in seeds.** *Plant Signal Behav* 6.5 (2011), 691–3 (see pp. 22, 23).
- [73] K. E. Shearwin, B. P. Callen and J. B. Egan. **Transcriptional interference - a crash course.** *Trends in Genetics* 21.6 (2005), 339–345 (see p. 22).
- [74] J. Camblong, N. Iglesias, C. Fickentscher, G. Dieppois and F. Stutz. **Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*.** *Cell* 131.4 (2007), 706–717 (see pp. 22, 34, 35).
- [75] M. A. Matzke and J. A. Birchler. **Rnai-mediated pathways in the nucleus.** *Nature Reviews Genetics* 6.1 (2005), 24–35 (see p. 22).
- [76] E. L. van Dijk, C. L. Chen, Y. d'Aubenton-Carafa, S. Gourvennec, M. Kwapisz, V. Roche, C. Bertrand, M. Silvain, P. Legoix-Né, S. Loeillet, A. Nicolas, C. Thermes and A. Morillon. **XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast.** *Nature* 475.7354 (2011), 114–117 (see pp. 22, 35, 36).
- [77] S. Geisler, L. Lojek, A. M. Khalil, K. E. Baker and J. Collier. **Article decapping of long noncoding RNAs regulates inducible genes.** *Molecular Cell* (2012), 1–13 (see p. 22).



- [78] Z. Xu, W. Wei, J. Gagneur, S. Clauder-Münster, M. Smolik, W. Huber and L. M. Steinmetz. **Antisense expression increases gene expression variability and locus interdependency.** *Molecular Systems Biology* 7.468 (2011), 468–468 (see pp. 22, 36).
- [79] R. L. Poole, G. L. A. Barker, K. Werner, G. F. Biggi, J. Coghill, J. G. Gibbings, S. Berry, J. M. Dunwell and K. J. Edwards. **Analysis of wheat sage tags reveals evidence for widespread antisense transcription.** *Bmc Genomics* 9.1 (2008), 475 (see p. 23).
- [80] F. Perocchi, Z. Y. Xu, S. Clauder-Munster and L. M. Steinmetz. **Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D.** *Nucleic Acids Research* 35.19 (2007), e128 (see p. 23).
- [81] J. M. Kuhn, G. Breton and J. I. Schroeder. **mRNA metabolism of flowering-time regulators in wild-type *Arabidopsis* revealed by a nuclear cap binding protein mutant, abh1.** *Plant Journal* 50.6 (2007), 1049–1062 (see p. 23).
- [82] S. P. Hazen, F. Naef, T. Quisel, J. M. Gendron, H. M. Chen, J. R. Ecker, J. O. Borevitz and S. A. Kay. **Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays.** *Genome biology* 10.2 (2009), R17 (see p. 23).
- [83] X. F. Zhou, R. Sunkar, H. L. Jin, J. K. Zhu and W. X. Zhang. **Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*.** *Genome Research* 19.1 (2009), 70–78 (see p. 23).
- [84] T. E. Coram, M. L. Settles and X. M. Chen. **Large-scale analysis of antisense transcription in wheat using the affymetrix genechip wheat genome array.** *BMC Genomics* 10.1 (2009), 253 (see p. 23).
- [85] F. Bardou, F. Merchan, F. Ariel and M. Crespi. **Dual RNAs in plants.** *Biochimie* 93.11 (2011), 1950–1954 (see p. 23).
- [86] Y. Kurihara, A. Matsui, K. Hanada, M. Kawashima, J. Ishida, T. Morosawa, M. Tanaka, E. Kaminuma, Y. Mochizuki, A. Matsushima, T. Toyoda, K. Shinozaki and M. Seki. **Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*.** *Proceedings of the National Academy of Sciences of the United States of America* 106.7 (2009), 2453–2458 (see p. 23).
- [87] C. H. Jen, I. Michalopoulos, D. R. Westhead and P. Meyer. **Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation.** *Genome biology* 6.1 (2005), R51 (see p. 23).
- [88] S. R. Henz, J. S. Cumbie, K. D. Kasschau, J. U. Lohmann, J. C. Carrington, D. Weigel and M. Schmid. **Distinct expression patterns of natural antisense transcripts in *Arabidopsis*.** *Plant Physiology* 144.3 (2007), 1247–1255 (see p. 23).
- [89] S. D. Michaels and R. M. Amasino. ***FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering.** *The Plant cell* 11.5 (1999), 949–56 (see pp. 24, 25, 133).

- [90] I. Baurle and C. Dean. **The timing of developmental transitions in plants.** *Cell* 125.4 (2006), 655–64 (see pp. 24, 25, 105, 133).
- [91] F. Fornara, A. de Montaigu and G. Coupland. **Snapshot: control of flowering in *Arabidopsis*.** *Cell* 141.3 (2010) (see pp. 24, 25).
- [92] W. W. Deng, H. Ying, C. A. Helliwell, J. M. Taylor, W. J. Peacock and E. S. Dennis. ***FLOWERING LOCUS C (FLC)* regulates development pathways throughout the life cycle of *Arabidopsis*.** *Proceedings of the National Academy of Sciences of the United States of America* 108.16 (2011), 6680–6685 (see pp. 24, 25).
- [93] S. Swiezewski, P. Crevillén, F. Q. Liu, J. R. Ecker, A. Jerzmanowski and C. Dean. **Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, *FLC*.** *Proceedings of the National Academy of Sciences of the United States of America* 104.9 (2007), 3633–3638 (see pp. 24, 25).
- [94] J. Liu, Y. H. He, R. Amasino and X. M. Chen. **siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*.** *Genes & Development* 18.23 (2004), 2873–2878 (see p. 24).
- [95] S. Swiezewski, F. Liu, A. Magusin and C. Dean. **Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target.** *Nature* 462.7274 (2009), 799–802 (see pp. 25, 30–32).
- [96] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. a. Brugmann, L. H. Goodnough, J. a. Helms, P. J. Farnham, E. Segal and H. Y. Chang. **Functional demarcation of active and silent chromatin domains in human *hox* loci by noncoding RNAs.** *Cell* 129.7 (2007), 1311–23 (see p. 25).
- [97] C. Hornyik, L. C. Terzi and G. G. Simpson. **The Spen family protein FPA controls alternative cleavage and polyadenylation of RNA.** *Developmental Cell* 18.2 (2010), 203–213 (see pp. 25, 29, 92, 134).
- [98] F. Liu, S. Marquardt, C. Lister, S. Swiezewski and C. Dean. **Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing.** *Science* 327.5961 (2010), 94–7 (see pp. 25, 27, 29, 47, 48, 51, 56, 83, 199).
- [99] S. C. Murray, A. S. Barros, D. A. Brown, P. Dudek, J. Ayling and J. Mellor. **A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription.** *Nucleic Acids Research* 40.6 (2012), 2432–2444 (see p. 25).
- [100] S. Oh, H. Zhang, P. Ludwig and S. van Nocker. **A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the *Arabidopsis FLC/MAF* MADS box gene family.** *The Plant Cell* 16.11 (2004), 2940–2953 (see p. 25).
- [101] Y. Cao, Y. Dai, S. J. Cui and L. G. Ma. **Histone H2B monoubiquitination in the chromatin of *FLOWERING LOCUS C* regulates flowering time in *Arabidopsis*.** *The Plant Cell* 20.10 (2008), 2586–2602 (see p. 25).

- [102] Z. Y. Xu, W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Munster, J. Camblong, E. Guffanti, F. Stutz, W. Huber and L. M. Steinmetz. **Bidirectional promoters generate pervasive transcription in yeast.** *Nature* 457.7232 (2009), 1033–U7 (see p. 25).
- [103] X. F. Gu, D. H. Jiang, Y. Q. Wang, A. Bachmair and Y. H. He. **Repression of the floral transition via histone H2B monoubiquitination.** *Plant Journal* 57.3 (2009), 522–533 (see p. 25).
- [104] S. Pien, D. Fleury, J. S. Mylne, P. Crevillén, D. Inze, Z. Avramova, C. Dean and U. Grossniklaus. **ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation.** *The Plant Cell* 20.3 (2008), 580–588 (see p. 25).
- [105] Y. Tamada, J.-Y. Yun, S. C. Woo and R. M. Amasino. **Arabidopsis trithorax-related7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of flowering locus c.** *The Plant cell* 21.10 (2009), 3257–69 (see p. 25).
- [106] D. H. Jiang, N. C. Kong, X. F. Gu, Z. C. Li and Y. H. He. **Arabidopsis compass-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development.** *PLoS Genetics* 7.3 (2011), e1001330 (see p. 25).
- [107] P. Crevillén and C. Dean. **Regulation of the floral repressor gene FLC: the complexity of transcription in a chromatin context.** *Current Opinion in Plant Biology* 14.1 (2011), 38–44 (see pp. 26, 95).
- [108] R. B. Deal, M. K. Kandasamy, E. C. McKinney and R. B. Meagher. **The nuclear actin-related protein ARP6 is a pleiotropic developmental regulator required for the maintenance of FLOWERING LOCUS C expression and repression of flowering in Arabidopsis.** *The Plant Cell* 17.10 (2005), 2633–2646 (see p. 26).
- [109] J. H. Clarke and C. Dean. **Mapping *Fri*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*.** *Molecular & General Genetics* 242.1 (1994), 81–89 (see p. 26).
- [110] U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino and C. Dean. **Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time.** *Science* 290.5490 (2000), 344–347 (see p. 26).
- [111] K. Choi, J. Kim, H.-J. Hwang, S. Kim, C. Park, S. Y. Kim and I. Lee. **The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors.** *The Plant Cell* 23.1 (2011), 289–303 (see pp. 26, 124, 135).
- [112] J. H. Ko, I. Mitina, Y. Tamada, Y. Hyun, Y. Choi, R. M. Amasino, B. Noh and Y. S. Noh. **Growth habit determination by the balance of histone methylation activities in Arabidopsis.** *EMBO Journal* 29.18 (2010), 3208–3215 (see pp. 26, 95, 124, 135).
- [113] H. C. Yang, M. Howard and C. Dean. **Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC.** *Current Biology* 24.15 (2014), 1793–1797 (see pp. 26, 81, 95).

- [114] R. Macknight, I. Bancroft, T. Page, C. Lister, R. Schmidt, K. Love, L. Westphal, G. Murphy, S. Sherson, C. Cobbett and C. Dean. **FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains.** *Cell* 89.5 (1997), 737–45 (see pp. 26, 72, 199).
- [115] G. G. Simpson, P. P. Dijkwel, V. Quesada, I. Henderson and C. Dean. **FY is an RNA 3' end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition.** *Cell* 113.6 (2003), 777–87 (see pp. 26, 83).
- [116] M. Ohnacker, S. M. L. Barabino, P. J. Preker and W. Keller. **The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3' end-processing complex.** *EMBO Journal* 19.1 (2000), 37–47 (see p. 26).
- [117] Y. S. Shi, D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, J. Frank and J. L. Manley. **Molecular architecture of the human pre-mRNA 3' processing complex.** *Molecular Cell* 33.3 (2009), 365–376 (see p. 26).
- [118] V. Quesada, R. Macknight, C. Dean and G. G. Simpson. **Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time.** *The EMBO journal* 22.12 (2003), 3142–52 (see p. 26).
- [119] C. Sonmez, I. Bäurle, A. Magusin, R. Dreos, S. Laubinger, D. Weigel and C. Dean. **RNA 3' processing functions of *Arabidopsis* FCA and FPA limit intergenic transcription.** *Proceedings of the National Academy of Sciences of the United States of America* 108.20 (2011), 8508–13 (see pp. 26, 27).
- [120] F. Liu, V. Quesada, P. Crevillén, I. Bäurle, S. Swiezewski and C. Dean. **The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*.** *Molecular Cell* 28.3 (2007), 398–407 (see pp. 27–29, 47, 79, 81, 83, 135, 199).
- [121] Y. J. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero and Y. Shi. **Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.** *Cell* 119.7 (2004), 941–953 (see p. 27).
- [122] W. J. Harris, X. Huang, J. T. Lynch, G. J. Spencer, J. R. Hitchin, Y. Y. Li, F. Ciceri, J. G. Blaser, B. F. Greystoke, A. M. Jordan, C. J. Miller, D. J. Ogilvie and T. C. P. Somervaille. **The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells.** *Cancer Cell* 21.4 (2012), 473–487 (see p. 27).
- [123] J. Wang, S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. T. Sun, H. Chang, G. L. Xu, F. Gaudet, E. Li and T. Chen. **The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation.** *Nature Genetics* 41.1 (2009), 125–129 (see p. 27).
- [124] X. H. Wu, M. Liu, B. Downie, C. Liang, G. L. Ji, Q. Q. Li and A. G. Hunt. **Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation.** *Proceedings of the National Academy of Sciences of the United States of America* 108.30 (2011), 12533–12538 (see p. 27).

- [125] I. R. Henderson, F. Q. Liu, S. Drea, G. G. Simpson and C. Dean. **An allelic series reveals essential roles for FY in plant development in addition to flowering-time control.** *Development* 132.16 (2005), 3597–3607 (see p. 27).
- [126] D. Manzano, S. Marquardt, A. M. E. Jones, I. Baurle, F. Q. Liu and C. Dean. **Altered interactions within FY/AtCPSF complexes required for *Arabidopsis* FCA-mediated chromatin silencing.** *Proceedings of the National Academy of Sciences of the United States of America* 106.21 (2009), 8772–8777 (see p. 27).
- [127] S. Marquardt, O. Raitskin, Z. Wu, F. Q. Liu, Q. W. Sun and C. Dean. **Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription.** *Molecular Cell* 54.1 (2014), 156–165 (see pp. 27, 48, 51, 56, 83, 92, 95, 134, 199, 201).
- [128] Z. W. Wang, Z. Wu, O. Raitskin, Q. W. Sun and C. Dean. **Antisense-mediated *FLC* transcriptional repression requires the P-TEFb transcription elongation factor.** *Proceedings of the National Academy of Sciences of the United States of America* 111.20 (2014), 7468–7473 (see pp. 27, 201).
- [129] Q. W. Sun, T. Csorba, K. Skourti-Stathaki, N. J. Proudfoot and C. Dean. **R-loop stabilization represses antisense transcription at the *Arabidopsis* *FLC* locus.** *Science* 340.6132 (2013), 619–621 (see p. 29).
- [130] H. Kamma, M. Fujimoto, M. Fujiwara, M. Matsui, H. Horiguchi, M. Hamasaki and H. Satoh. **Interaction of hnRNP A2/B1 isoforms with telomeric ssDNA and the in vitro function.** *Biochemical and Biophysical Research Communications* 280.3 (2001), 625–630 (see p. 29).
- [131] I. Baurle and C. Dean. **Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of *Arabidopsis* targets.** *Plos ONE* 3.7 (2008), e2733 (see p. 29).
- [132] E. Rosonina and J. L. Manley. **Alternative polyadenylation blooms.** *Developmental Cell* 18.2 (2010), 172–174 (see p. 29).
- [133] C. C. Sheldon, D. T. Rouse, E. J. Finnegan, W. J. Peacock and E. S. Dennis. **The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*).** *Proceedings of the National Academy of Sciences of the United States of America* 97.7 (2000), 3753–3758 (see p. 29).
- [134] F. De Lucia, P. Crevillén, A. M. E. Jones, T. Greb and C. Dean. **A PHD-Polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization.** *Proceedings of the National Academy of Sciences of the United States of America* 105.44 (2008), 16831–6 (see pp. 29–31).
- [135] C. C. Sheldon, A. B. Conn, E. S. Dennis and W. J. Peacock. **Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression.** *The Plant Cell* 14.October (2002), 2527–2537 (see p. 29).

- [136] C. C. Wood, M. Robertson, G. Tanner, W. J. Peacock, E. S. Dennis and C. A. Helliwell. **The *Arabidopsis thaliana* vernalization response requires a Polycomb-like protein complex that also includes *VERNALIZATION INSENSITIVE 3***. *Proceedings of the National Academy of Sciences of the United States of America* 103.39 (2006), 14631–14636 (see p. 30).
- [137] E. J. Finnegan and E. S. Dennis. **Vernalization-induced trimethylation of histone H3 lysine 27 at *FLC* is not maintained in mitotically quiescent cells**. *Current Biology* 17.22 (2007), 1978–83 (see p. 30).
- [138] A. Angel, J. Song, C. Dean and M. Howard. **A Polycomb-based switch underlying quantitative epigenetic memory**. *Nature* 476.7358 (2011), 105–108 (see pp. 30, 31, 81, 103, 133).
- [139] A. Angel, J. Song, H. Yang, J. I. Questa, C. Dean and M. Howard. **Vernalizing cold is registered digitally at *FLC***. *Proceedings of the National Academy of Sciences of the United States of America* 112.13 (2015), 4146–4151 (see pp. 30, 133).
- [140] J. B. Heo and S. Sung. **Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA**. *Science* 331.6013 (2011), 76–9 (see pp. 31, 32).
- [141] C. a. Helliwell, M. Robertson, E. J. Finnegan, D. M. Buzas and E. S. Dennis. **Vernalization-repression of *Arabidopsis FLC* requires promoter sequences but not antisense transcripts**. *PLoS ONE* 6.6 (2011), e21513–e21513 (see p. 32).
- [142] T. Csorba, J. I. Questa, Q. W. Sun and C. Dean. **Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization**. *Proceedings of the National Academy of Sciences of the United States of America* 111.45 (2014), 16160–16165 (see p. 32).
- [143] F. Mohammad, T. Mondal and C. Kanduri. **Epigenetics of imprinted long noncoding RNAs**. *Epigenetics* 4.5 (2009), 277–286 (see pp. 33, 36).
- [144] S. L. Bumgarner, R. D. Dowell, P. Grisafi, D. K. Gifford and G. R. Fink. **Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast**. *Proceedings of the National Academy of Sciences of the United States of America* 106.43 (2009), 18321–18326 (see pp. 33, 105–107).
- [145] S. L. Bumgarner, G. Neuert, B. F. Voight, A. Symbor-nagrabaska, P. Grisafi, A. V. Oudenaarden and G. R. Fink. **Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment**. *Molecular Cell* (2012), 1–13 (see pp. 33, 34, 201).
- [146] J. T. Lee. **Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control**. *Nature Reviews Molecular Cell Biology* 12.12 (2011), 815–826 (see pp. 34, 36).
- [147] S. Augui, E. P. Nora and E. Heard. **Regulation of X-chromosome inactivation by the X-inactivation centre**. *Nature Reviews Genetics* 12.6 (2011), 429–442 (see pp. 34, 36).



- [148] J. Camblong, N. Beyrouthy, E. Guffanti, G. Schlaepfer, L. M. Steinmetz and F. Stutz. **Trans-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*.** *Genes & Development* 23.13 (2009), 1534–1545 (see p. 35).
- [149] M. Tisseur, M. Kwapisz and A. Morillon. **Pervasive transcription - lessons from yeast.** *Biochimie* 93.11 (2011), 1889–1896 (see p. 35).
- [150] J. Berretta, M. Pinskaya and A. Morillon. **A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*.** *Genes & Development* 22.5 (2008), 615–626 (see p. 35).
- [151] L. Redrup, M. R. Branco, E. R. Perdeaux, C. Krueger, A. Lewis, F. Santos, T. Nagano, B. S. Cobb, P. Fraser and W. Reik. **The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing.** *Development* 136.4 (2009), 525–530 (see p. 36).
- [152] R. R. Pandey, T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-DiNardo and C. Kanduri. ***Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.** *Molecular Cell* 32.2 (2008), 232–246 (see p. 36).
- [153] J. H. R. Ietswaart. **Modelling the segregation mechanism of low copy number plasmid pB171.** Masters thesis. 2011 (see pp. 37, 38, 202, 203).
- [154] R. Reyes-Lamothe, T. Tran, D. Meas, L. Lee, A. M. Li, D. J. Sherratt and M. E. Tolmasky. **High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division.** *Nucleic Acids Research* 42.2 (2014), 1042–1051 (see p. 38).
- [155] K. Gerdes, M. Howard and F. Szardenings. **Pushing and pulling in prokaryotic DNA segregation.** *Cell* 141.6 (2010), 927–942 (see pp. 38, 182, 202).
- [156] G. Ebersbach and K. Gerdes. **The double par locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA.** *Proceedings of the National Academy of Sciences of the United States of America* 98.26 (2001), 15078–15083 (see pp. 38, 39, 144, 167, 174, 203).
- [157] G. Ebersbach and K. Gerdes. **Bacterial mitosis: partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell.** *Molecular Microbiology* 52.2 (2004), 385–398 (see pp. 38–40, 44, 150, 167).
- [158] G. Ebersbach, S. Ringgaard, J. Moller-Jensen, Q. Wang, D. J. Sherratt and K. Gerdes. **Regular cellular distribution of plasmids by oscillating and filament-forming ParA ATPase of plasmid pB171.** *Molecular Microbiology* 61.6 (2006), 1428–1442 (see pp. 38–40, 42, 150, 167, 204).

- [159] S. Ringgaard, J. van Zon, M. Howard and K. Gerdes. **Movement and equipositioning of plasmids by ParA filament disassembly.** *Proceedings of the National Academy of Sciences of the United States of America* 106.46 (2009), 19369–19374 (see pp. 38–40, 42, 43, 144, 150, 154, 155, 167, 170, 174, 185, 186, 202–204).
- [160] A. Derome, C. Hoischen, M. Bussiek, R. Grady, M. Adamczyk, B. Kedzierska, S. Diekmann, D. Barilla and F. Hayes. **Centromere anatomy in the multidrug-resistant pathogen enterococcus faecium.** *Proceedings of the National Academy of Sciences of the United States of America* 105.6 (2008), 2151–2156 (see pp. 38, 39).
- [161] J. K. Polka, J. M. Kollman and R. D. Mullins. **Accessory factors promote Alfa-dependent plasmid segregation by regulating filament nucleation, disassembly, and bundling.** *Proceedings of the National Academy of Sciences of the United States of America* 111.6 (2014), 2176–2181 (see pp. 38, 158, 159, 185).
- [162] J. Moller-Jensen, R. B. Jensen, J. Lowe and K. Gerdes. **Prokaryotic DNA segregation by an actin-like filament.** *Embo Journal* 21.12 (2002), 3119–3127 (see p. 38).
- [163] C. S. Campbell and R. D. Mullins. **In vivo visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids.** *Journal of Cell Biology* 179.5 (2007), 1059–1066 (see p. 38).
- [164] P. Gayathri, T. Fujii, J. Moller-Jensen, F. van den Ent, K. Namba and J. Lowe. **A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation.** *Science* 338.6112 (2012), 1334–1337 (see p. 38).
- [165] R. A. Larsen, C. Cusumano, A. Fujioka, G. Lim-Fong, P. Patterson and J. Pogliano. **Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in bacillus thuringiensis.** *Genes & Development* 21.11 (2007), 1340–1352 (see p. 38).
- [166] A. G. Vecchiarelli, Y. W. Han, X. Tan, M. Mizuuchi, R. Ghirlando, C. Biertumpfel, B. E. Funnell and K. Mizuuchi. **ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition.** *Molecular Microbiology* 78.1 (2010), 78–91 (see pp. 38–40, 42, 43, 146, 153, 154, 182, 184, 185, 198, 202).
- [167] J. L. Ptacin, S. F. Lee, E. C. Garner, E. Toro, M. Eckart, L. R. Comolli, W. Moerner and L. Shapiro. **A spindle-like apparatus guides bacterial chromosome segregation.** *Nature Cell Biology* 12.8 (2010), 791–U46 (see pp. 38–40, 42, 186).
- [168] M. A. Fogel and M. K. Waldor. **A dynamic, mitotic-like mechanism for bacterial chromosome segregation.** *Genes & Development* 20.23 (2006), 3269–3282 (see pp. 38, 40).
- [169] H. C. Lim, I. V. Surovtsev, B. G. Beltran, F. Huang, J. Bewersdorf and C. Jacobs-Wagner. **Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation.** *Elife* 3 (2014) (see pp. 38, 40, 42, 186).
- [170] S. Ringgaard, J. Lowe and K. Gerdes. **Centromere pairing by a plasmid-encoded type I parB protein.** *Journal of Biological Chemistry* 282.38 (2007), 28216–28225 (see pp. 39, 144).



- [171] S. Ringgaard, G. Ebersbach, J. Borch and K. Gerdes. **Regulatory cross-talk in the double par locus of plasmid pB171.** *Journal of Biological Chemistry* 282.5 (2007), 3134–3145 (see pp. 39, 144, 151).
- [172] O. Rodionov and M. Yarmolinsky. **Plasmid partitioning and the spreading of P1 partition protein parB.** *Molecular Microbiology* 52.4 (2004), 1215–1223 (see p. 39).
- [173] T. Hatano and H. Niki. **Partitioning of P1 plasmids by gradual distribution of the ATPase ParA.** *Molecular Microbiology* 78.5 (2010), 1182–1198 (see pp. 39, 42, 43, 146, 172, 182, 184, 202).
- [174] M. A. J. Roberts, G. H. Wadhams, K. A. Hadfield, S. Tickner and J. P. Armitage. **ParA-like protein uses nonspecific chromosomal DNA binding to partition protein complexes.** *Proceedings of the National Academy of Sciences of the United States of America* 109.17 (2012), 6698–6703 (see pp. 39, 43).
- [175] T. Hatano, Y. Yamaichi and H. Niki. **Oscillating focus of SopA associated with filamentous structure guides partitioning of F plasmid.** *Molecular Microbiology* 64.5 (2007), 1198–1213 (see pp. 40, 202).
- [176] A. G. Vecchiarelli, K. C. Neuman and K. Mizuuchi. **A propagating ATPase gradient drives transport of surface-confined cellular cargo.** *Proceedings of the National Academy of Sciences of the United States of America* 111.13 (2014), 4880–4885 (see pp. 40, 42, 146, 182, 184).
- [177] L. C. Hwang, A. G. Vecchiarelli, Y. W. Han, M. Mizuuchi, Y. Harada, B. E. Funnell and K. Mizuuchi. **ParA-mediated plasmid partition driven by protein pattern self-organization.** *EMBO Journal* 32.9 (2013), 1238–1249 (see pp. 40, 42, 146, 150, 182, 184).
- [178] D. Barilla, M. F. Rosenberg, U. Nobbmann and F. Hayes. **Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF.** *Embo Journal* 24.7 (2005), 1453–1464 (see p. 40).
- [179] G. E. Lim, A. I. Derman and J. Pogliano. **Bacterial DNA segregation by dynamic SopA polymers.** *Proceedings of the National Academy of Sciences of the United States of America* 102.49 (2005), 17658–17663 (see p. 40).
- [180] A. G. Vecchiarelli, L. C. Hwang and K. Mizuuchi. **Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism.** *Proceedings of the National Academy of Sciences of the United States of America* 110.15 (2013), E1390–E1397 (see pp. 40, 42, 146, 150, 182, 184, 185, 198).
- [181] A. G. Vecchiarelli, K. Mizuuchi and B. E. Funnell. **Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria.** *Molecular Microbiology* 86.3 (2012), 513–523 (see p. 40).
- [182] F. Szardenings, D. Guymer and K. Gerdes. **ParA ATPases can move and position DNA and subcellular structures.** *Current Opinion in Microbiology* 14.6 (2011), 712–718 (see p. 41).

- [183] M. Sengupta, H. J. Nielsen, B. Youngren and S. Austin. **P1 plasmid segregation: accurate redistribution by dynamic plasmid pairing and separation.** *Journal of Bacteriology* 192.5 (2010), 1175–1183 (see pp. 40, 42, 154, 172).
- [184] M. Howard and K. Gerdes. **What is the mechanism of ParA-mediated DNA movement?** *Molecular Microbiology* 78.1 (2010), 9–12 (see pp. 40, 202).
- [185] E. J. Banigan, M. A. Gelbart, Z. Gitai, N. S. Wingreen and A. J. Liu. **Filament depolymerization can explain chromosome pulling during bacterial mitosis.** *PLoS Computational Biology* 7.9 (2011), e1002145 (see pp. 42, 186).
- [186] T. Sugawara and K. Kaneko. **Chemophoresis as a driving force for intracellular organization: theory and application to plasmid partitioning.** *Biophysics - The Biophysical Society of Japan* 7 (2011), 77–88 (see pp. 42, 146, 182, 184).
- [187] D. F. Savage, B. Afonso, A. H. Chen and P. A. Silver. **Spatially ordered dynamics of the bacterial carbon fixation machinery.** *Science* 327.5970 (2010), 1258–1261 (see p. 43).
- [188] S. R. Thompson, G. H. Wadhams and J. P. Armitage. **The positioning of cytoplasmic protein clusters in bacteria.** *Proceedings of the National Academy of Sciences of the United States of America* 103.21 (2006), 8209–8214 (see p. 43).
- [189] J. K. Fisher, A. Bourniquel, G. Witz, B. Weiner, M. Prentiss and N. Kleckner. **Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells.** *Cell* 153.4 (2013), 882–895 (see pp. 43, 44, 170, 187).
- [190] N. Hadizadeh Yazdi, C. C. Guet, R. C. Johnson and J. F. Marko. **Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions.** *Molecular Microbiology* 86.6 (2012), 1318–33 (see pp. 43, 170, 187).
- [191] P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner and J. Kondev. **Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament.** *Proceedings of the National Academy of Sciences of the United States of America* 107.11 (2010), 4991–4995 (see pp. 43, 170).
- [192] M. A. Umbarger, E. Toro, M. A. Wright, G. J. Porreca, D. Bau, S. H. Hong, M. J. Fero, L. J. Zhu, M. A. Marti-Renom, H. H. McAdams, L. Shapiro, J. Dekker and G. M. Church. **The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation.** *Molecular Cell* 44.2 (2011), 252–264 (see p. 44).
- [193] I. A. Berlatzky, A. Rouvinski and S. Ben-Yehuda. **Spatial organization of a replicating bacterial chromosome.** *Proceedings of the National Academy of Sciences of the United States of America* 105.37 (2008), 14136–14140 (see p. 44).
- [194] R. Reyes-Lamothe, E. Nicolas and D. J. Sherratt. **Chromosome replication and segregation in bacteria.** *Annual Review of Genetics* 46 (2012), 121–143 (see p. 44).
- [195] R. Mercier, M. A. Petit, S. Schbath, S. Robin, M. El Karoui, F. Boccard and O. Espeli. **The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain.** *Cell* 135.3 (2008), 475–485 (see pp. 44, 174).

- [196] A. Badrinarayanan, C. Lesterlin, R. Reyes-Lamothe and D. Sherratt. **The *Escherichia coli* SMC complex, MukBEE, shapes nucleoid organization independently of DNA replication.** *Journal of Bacteriology* 194.17 (2012), 4669–4676 (see pp. 44, 174).
- [197] B. Ezaki, T. Ogura, H. Niki and S. Hiraga. **Partitioning of a mini-F plasmid into anucleate cells of the Mukb null mutant.** *Journal of Bacteriology* 173.20 (1991), 6643–6646 (see pp. 44, 174).
- [198] B. E. Funnell and L. Gagnier. **Partition of P1 plasmids in *Escherichia coli* mukb chromosomal partition mutants.** *Journal of Bacteriology* 177.9 (1995), 2381–2386 (see pp. 44, 174).
- [199] B. R. Parry, I. V. Surovtsev, M. T. Cabeen, C. S. O’Hem, E. R. Dufresne and C. Jacobs-Wagner. **The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity.** *Cell* 156.1-2 (2014), 183–194 (see pp. 44, 156, 158, 159, 185).
- [200] A. I. Derman, G. Lim-Fong and J. Pogliano. **Intracellular mobility of plasmid DNA is limited by the ParA family of partitioning systems.** *Molecular Microbiology* 67.5 (2008), 935–946 (see pp. 44, 185).
- [201] M. L. Chou and C. H. Yang. **Late-flowering genes interact with early-flowering genes to regulate flowering time in *Arabidopsis thaliana*.** *Plant & Cell Physiology* 40.7 (1999), 702–708 (see p. 48).
- [202] T. Page, R. Macknight, C. H. Yang and C. Dean. **Genetic interactions of the *Arabidopsis* flowering time gene FCA, with genes regulating floral initiation.** *Plant Journal* 17.3 (1999), 231–239 (see p. 48).
- [203] J. M. Alonso et al. **Genome-wide insertional mutagenesis of *Arabidopsis thaliana*.** *Science* 301.5633 (2003), 653–657 (see p. 48).
- [204] A. H. Ehrensberger, G. P. Kelly and J. Q. Svejstrup. **Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps.** *Cell* 154.4 (2013), 713–715 (see pp. 51–53, 71, 93).
- [205] C. G. Danko, N. Hah, X. Luo, A. L. Martins, L. Core, J. T. Lis, A. Siepel and W. L. Kraus. **Signaling pathways differentially affect RNA Polymerase II initiation, pausing, and elongation rate in cells.** *Molecular Cell* 50.2 (2013), 212–222 (see pp. 52, 62, 93, 101).
- [206] I. Jonkers, H. Kwak and J. T. Lis. **Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons.** *Elife* 3 (2014) (see pp. 52, 95).
- [207] A. C. M. Cheung and P. Cramer. **A movie of RNA Polymerase II transcription.** *Cell* 149.7 (2012), 1431–1437 (see p. 52).
- [208] Z. Wu, R. Ietswaart, F. Liu, H. Yang, M. Howard and C. Dean. **Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation.** *Proc Natl Acad Sci U S A* 113.1 (2016), 218–23 (see pp. 54, 72, 73, 78–80).

- [209] G. Fuchs, Y. Voichek, S. Benjamin, S. Gilad, I. Amit and M. Oren. **4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells.** *Genome Biology* 15.5 (2014), R69 (see p. 62).
- [210] J. Singh and R. A. Padgett. **Rates of in situ transcription and splicing in large human genes.** *Nature Structural & Molecular Biology* 16.11 (2009), 1128–U1 (see p. 62).
- [211] P. Maiuri, A. Knezevich, A. De Marco, D. Mazza, A. Kula, J. G. McNally and A. Marcelllo. **Fast transcription rates of RNA Polymerase II in human cells.** *Embo Reports* 12.12 (2011), 1280–1285 (see pp. 62, 93).
- [212] D. M. Bhatt, A. Pandya-Jones, A. J. Tong, I. Barozzi, M. M. Lissner, G. Natoli, D. L. Black and S. T. Smale. **Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions.** *Cell* 150.2 (2012), 279–290 (see pp. 62, 64, 68, 88).
- [213] J. M. Gray, D. A. Harmin, S. A. Boswell, N. Cloonan, T. E. Mullen, J. J. Ling, N. Miller, S. Kuersten, Y. C. Ma, S. A. McCarroll, S. M. Grimmond and M. Springer. **Snapshot-seq: a method for extracting genome-wide, *in vivo* mRNA dynamics from a single total RNA sample.** *Plos ONE* 9.2 (2014), e89673 (see pp. 62, 63, 90).
- [214] J. R. Hesselberth. **Lives that introns lead after splicing.** *Wiley Interdisciplinary Reviews-Rna* 4.6 (2013), 677–691 (see pp. 63, 69, 130).
- [215] P. B. Mason and K. Struhl. **Distinction and relationship between elongation rate and processivity of RNA Polymerase II *in vivo*.** *Molecular Cell* 17.6 (2005), 831–40 (see pp. 69, 70).
- [216] L. Weinberger, Y. Voichek, I. Tirosh, G. Hornung, I. Amit and N. Barkai. **Expression noise and acetylation profiles distinguish HDAC functions.** *Molecular Cell* 47.2 (2012), 193–202 (see p. 70).
- [217] I. Gy, V. Gascioli, D. Lauressergues, J. B. Morel, J. Gombert, F. Proux, C. Proux, H. Vaucheret and A. C. Mallory. ***Arabidopsis* FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors.** *The Plant Cell* 19.11 (2007), 3451–3461 (see p. 78).
- [218] A. Slepoy, A. P. Thompson and S. J. Plimpton. **A constant-time kinetic monte carlo algorithm for simulation of large biochemical reaction networks.** *Journal of Chemical Physics* 128.20 (2008), 205101 (see pp. 79, 115, 127, 151).
- [219] S. Aitken, R. D. Alexander and J. D. Beggs. **Modelling reveals kinetic advantages of co-transcriptional splicing.** *PLoS Computational Biology* 7.10 (2011), e1002215–e1002215 (see pp. 103, 130).
- [220] R. B. Deal, J. G. Henikoff and S. Henikoff. **Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones.** *Science* 328.5982 (2010), 1161–1164 (see pp. 103, 135).
- [221] L. J. Core, J. J. Waterfall, D. A. Gilchrist, D. C. Fargo, H. Kwak, K. Adelman and J. T. Lis. **Defining the status of RNA polymerase at promoters.** *Cell Reports* 2.4 (2012), 1025–1035 (see p. 93).

- [222] M. Voliotis, N. Cohen, C. Molina-Paris and T. B. Liverpool. **Fluctuations, pauses, and backtracking in DNA transcription.** *Biophysical Journal* 94.2 (2008), 334–348 (see pp. 95, 106, 108, 133).
- [223] M. Saponaro, T. Kantidakis, R. Mitter, G. P. Kelly, M. Heron, H. Williams, J. Soding, A. Stewart and J. Q. Svejstrup. **RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress.** *Cell* 157.5 (2014), 1037–1049 (see p. 95).
- [224] C. M. Weber, S. Ramachandran and S. Henikoff. **Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase.** *Molecular Cell* 53.5 (2014), 819–830 (see p. 95).
- [225] S. Y. Kim, Y. H. He, Y. Jacob, Y. S. Noh, S. Michaels and R. Amasino. **Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyl transferase.** *The Plant Cell* 17.12 (2005), 3301–3310 (see p. 95).
- [226] Y. H. He, S. D. Michaels and R. M. Amasino. **Regulation of flowering time by histone acetylation in *Arabidopsis*.** *Science* 302.5651 (2003), 1751–1754 (see p. 95).
- [227] X. H. Yu and S. D. Michaels. **The *Arabidopsis* Paf1c complex component CDC73 participates in the modification of *FLOWERING LOCUS C* chromatin.** *Plant Physiology* 153.3 (2010), 1074–1084 (see p. 95).
- [228] M. K. Gardner, B. L. Sprague, C. G. Pearson, B. D. Cosgrove, A. D. Bicek, K. Bloom, E. D. Salmon and D. J. Odde. **Model convolution: a computational approach to digital image interpretation.** *Cellular and Molecular Bioengineering* 3.2 (2010), 163–170 (see p. 98).
- [229] J. C. Pearson, D. Lemons and W. McGinnis. **Modulating hox gene functions during animal body patterning.** *Nature Reviews Genetics* 6.12 (2005), 893–904 (see p. 105).
- [230] K. E. Jaeger, N. Pullen, S. Lamzin, R. J. Morris and P. A. Wigge. **Interlocking feedback loops govern the dynamic behavior of the floral transition in *Arabidopsis*.** *The Plant Cell* 25.3 (2013), 820–833 (see p. 105).
- [231] L. Corbesier, C. Vincent, S. H. Jang, F. Fornara, Q. Z. Fan, I. Searle, A. Giakountis, S. Farrona, L. Gissot, C. Turnbull and G. Coupland. **FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*.** *Science* 316.5827 (2007), 1030–1033 (see pp. 105, 134, 201).
- [232] K. E. Jaeger and P. A. Wigge. **FT protein acts as a long-range signal in *Arabidopsis*.** *Current Biology* 17.12 (2007), 1050–1054 (see pp. 105, 134, 201).
- [233] J. Mathieu, N. Warthmann, F. Kuttner and M. Schmid. **Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*.** *Current Biology* 17.12 (2007), 1055–1060 (see pp. 105, 134, 201).
- [234] A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden and S. Tyagi. **Imaging individual mRNA molecules using multiple singly labeled probes.** *Nature Methods* 5.10 (2008), 877–879 (see pp. 105, 133).

- [235] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh and A. Raj. **Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms.** *Molecular Cell* 58.2 (2015), 339–352 (see pp. 105, 107, 116–118, 133, 135).
- [236] M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain. **Stochastic gene expression in a single cell.** *Science* 297.5584 (2002), 1183–1186 (see p. 105).
- [237] P. S. Swain, M. B. Elowitz and E. D. Siggia. **Intrinsic and extrinsic contributions to stochasticity in gene expression.** *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (2002), 12795–12800 (see pp. 105, 106).
- [238] I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox. **Real-time kinetics of gene activity in individual bacteria.** *Cell* 123.6 (2005), 1025–1036 (see pp. 105, 108, 133).
- [239] D. Zenklusen, D. R. Larson and R. H. Singer. **Single-RNA counting reveals alternative modes of gene expression in yeast.** *Nature Structural & Molecular Biology* 15.12 (2008), 1263–1271 (see pp. 105, 106, 108, 118, 133).
- [240] K. B. Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg and S. Itzkovitz. **Bursty gene expression in the intact mammalian liver.** *Molecular Cell* 58.1 (2015), 147–156 (see pp. 105, 108, 118, 133).
- [241] J. C. W. Locke, J. W. Young, M. Fontes, M. J. H. Jimenez and M. B. Elowitz. **Stochastic pulse regulation in bacterial stress response.** *Science* 334.6054 (2011), 366–369 (see p. 106).
- [242] A. Raj and A. van Oudenaarden. **Nature, nurture, or chance: stochastic gene expression and its consequences.** *Cell* 135.2 (2008), 216–226 (see p. 107).
- [243] S. Duncan. **Predicting the impact of climate change on vernalization for *Arabidopsis thaliana*.** PhD thesis. 2015 (see pp. 107, 109, 113, 118).
- [244] C. W. Gardiner. **Handbook of stochastic methods.** Springer, 2009 (see pp. 108, 110, 135, 136, 138).
- [245] V. Shahrezaei and P. S. Swain. **Analytical distributions for stochastic gene expression.** *Proceedings of the National Academy of Sciences of the United States of America* 105.45 (2008), 17256–17261 (see pp. 108, 136, 140).
- [246] J. R. Chubb and T. B. Liverpool. **Bursts and pulses: insights from single cell studies into transcriptional mechanisms.** *Current Opinion in Genetics & Development* 20.5 (2010), 478–484 (see p. 108).
- [247] L. Dolken, Z. Ruzsics, B. Radle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal and U. H. Koszinowski. **High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay.** *RNA* 14.9 (2008), 1959–1972 (see pp. 110, 111).
- [248] K. Sidaway-Lee, M. J. Costa, D. A. Rand, B. Finkenstadt and S. Penfield. **Direct measurement of transcription rates reveals multiple mechanisms for configuration of the *Arabidopsis* ambient temperature response.** *Genome Biology* 15.3 (2014), R45 (see pp. 110, 134).



- [249] R. Narsai, K. a. Howell, a. H. Millar, N. O'Toole, I. Small and J. Whelan. **Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*.** *The Plant cell* 19.11 (2007), 3418–36 (see p. 111).
- [250] K. Yin, M. Ueda, H. Takagi, T. Kajihara, S. S. Aki, T. Nobusawa, C. Umeda-Hara and M. Umeda. **A dual-color marker system for *in vivo* visualization of cell cycle progression in *Arabidopsis*.** *Plant Journal* 80.3 (2014), 541–552 (see pp. 113, 115, 140).
- [251] K. Hayashi, J. Hasegawa and S. Matsunaga. **The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion.** *Scientific Reports* 3 (2013), 2723 (see pp. 113, 115, 140).
- [252] B. B. Buchanan, W. Gruissem and R. L. Jones. **Biochemistry & molecular biology of plants.** American Society of Plant Physiologists, 2000 (see p. 116).
- [253] T. Trcek, J. A. Chao, D. R. Larson, H. Y. Park, D. Zenklusen, S. M. Shenoy and R. H. Singer. **Single-mRNA counting using fluorescent *in situ* hybridization in budding yeast.** *Nature Protocols* 7.2 (2012), 408–419 (see p. 118).
- [254] R. E. Thompson, D. R. Larson and W. W. Webb. **Precise nanometer localization analysis for individual fluorescent probes.** *Biophysical Journal* 82.5 (2002), 2775–2783 (see p. 118).
- [255] S. Rosa, F. De Lucia, J. S. Mylne, D. Zhu, N. Ohmido, A. Pendle, N. Kato, P. Shaw and C. Dean. **Physical clustering of *FLC* alleles during Polycomb-mediated epigenetic silencing in vernalization.** *Genes & Development* 27.17 (2013), 1845–1850 (see p. 119).
- [256] P. J. Li, Z. Tao and C. Dean. **Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*.** *Genes & Development* 29.7 (2015), 696–701 (see pp. 124, 126, 127, 134, 201).
- [257] N. Geraldo, I. Baurle, S. Kidou, X. Y. Hu and C. Dean. **FRIGIDA delays flowering in *Arabidopsis* via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex.** *Plant Physiology* 150.3 (2009), 1611–1618 (see p. 124).
- [258] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao and R. H. Singer. **Real-time observation of transcription initiation and elongation on an endogenous yeast gene.** *Science* 332.6028 (2011), 475–478 (see p. 133).
- [259] S. Berry, M. Hartley, T. S. Olsson, C. Dean and M. Howard. **Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance.** *Elife* 4 (2015) (see p. 133).
- [260] C. C. Sheldon, J. E. Burn, P. P. Perez, J. Metzger, J. A. Edwards, W. J. Peacock and E. S. Dennis. **The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation.** *The Plant Cell* 11.3 (1999), 445–458 (see p. 133).
- [261] S. D. Michaels and R. M. Amasino. **Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization.** *The Plant Cell* 13.4 (2001), 935–941 (see p. 133).

- [262] P. J. Li, D. Filiault, M. S. Box, E. Kerdaffrec, C. van Oosterhout, A. M. Wilczek, J. Schmitt, M. McMullan, J. Bergelson, M. Nordborg and C. Dean. **Multiple *FLC* haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*.** *Genes & Development* 28.15 (2014), 1635–1640 (see p. 133).
- [263] V. Coustham, P. J. Li, A. Strange, C. Lister, J. Song and C. Dean. **Quantitative modulation of Polycomb silencing underlies natural variation in vernalization.** *Science* 337.6094 (2012), 584–587 (see p. 133).
- [264] A. Strange, P. Li, C. Lister, J. Anderson, N. Warthmann, C. Shindo, J. Irwin, M. Nordborg and C. Dean. **Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions.** *PLoS ONE* 6.5 (2011), e19949 (see p. 133).
- [265] C. Shindo, C. Lister, P. Crevillén, M. Nordborg and C. Dean. **Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response.** *Genes & Development* 20.22 (2006), 3079–3083 (see p. 133).
- [266] K. D. Edwards, P. E. Anderson, A. Hall, N. S. Salathia, J. C. W. Locke, J. R. Lynn, M. Straume, J. Q. Smith and A. J. Millar. ***FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock.** *The Plant Cell* 18.3 (2006), 639–650 (see p. 133).
- [267] M. Rabani, R. Raychowdhury, M. Jovanovic, M. Rooney, D. J. Stumpo, A. Pauli, N. Hacohen, A. F. Schier, P. J. Blackshear, N. Friedman, I. Amit and A. Regev. **High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies.** *Cell* 159.7 (2014), 1698–1710 (see p. 134).
- [268] D. J. Katz, T. M. Edwards, V. Reinke and W. G. Kelly. **A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory.** *Cell* 137.2 (2009), 308–320 (see p. 135).
- [269] L. J. Gaydos, W. C. Wang and S. Strome. **H3K27me and PRC2 transmit a memory of repression across generations and during development.** *Science* 345.6203 (2014), 1515–1518 (see p. 135).
- [270] D. Zwillinger. **Handbook of differential equations.** 3rd. Academic Press, 1998 (see p. 136).
- [271] R. Ietswaart, F. Szardenings, K. Gerdes and M. Howard. **Competing ParA structures space bacterial plasmids equally over the nucleoid.** *PLoS Computational Biology* 10.12 (2014), e1004009 (see pp. 144, 154, 193, 203).
- [272] O. Sliusarenko, J. Heinritz, T. Emonet and C. Jacobs-Wagner. **High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics.** *Molecular Microbiology* 80.3 (2011), 612–627 (see pp. 144, 146, 168, 169, 171).
- [273] S. C. Weber, A. J. Spakowitz and J. A. Theriot. **Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm.** *Physical Review Letters* 104.23 (2010), 238102 (see p. 158).



- [274] S. C. Weber, M. A. Thompson, W. E. Moerner, A. J. Spakowitz and J. A. Theriot. **Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function.** *Biophysical Journal* 102.11 (2012), 2443–2450 (see p. 158).
- [275] M. T. Swulius and G. J. Jensen. **The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-terminal yellow fluorescent protein tag.** *Journal of Bacteriology* 194.23 (2012), 6382–6386 (see p. 174).
- [276] M. Nollmann, N. J. Crisona and P. B. Arimondo. **Thirty years of *Escherichia coli* DNA gyrase: from *in vivo* function to single-molecule mechanism.** *Biochimie* 89.4 (2007), 490–499 (see p. 174).
- [277] W. T. Tao, S. Dasgupta and K. Nordstrom. **Role of the *mukB* gene in chromosome and plasmid partition in *Escherichia coli*.** *Molecular Microbiology* 38.2 (2000), 392–400 (see p. 177).
- [278] E. M. M. Manders, F. J. Verbeek and J. A. Aten. **Measurement of colocalization of objects in dual-color confocal images.** *Journal of Microscopy-Oxford* 169 (1993), 375–382 (see pp. 180, 196).
- [279] P. Thevenaz, U. E. Ruttimann and M. Unser. **A pyramid approach to subpixel registration based on intensity.** *IEEE Transactions on Image Processing* 7.1 (1998), 27–41 (see p. 196).
- [280] A. R. Gendall, Y. Y. Levy, A. Wilson and C. Dean. **The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*.** *Cell* 107.4 (2001), 525–535 (see p. 202).



## APPENDIX: PUBLICATIONS ASSOCIATED TO THIS THESIS

---

R. Ietswaart, Z. Wu, C. Dean. **Flowering time control: another window to the connection between antisense RNA and chromatin.** *Trends Genet.* 28: 445-53 (2012)

R. Ietswaart, F. Szardenings, K. Gerdes, M. Howard. **Competing ParA structures space bacterial plasmids equally over the nucleoid.** *PLoS Comput Biol.* 10(12): e1004009 (2014)

Z. Wu, R. Ietswaart, F. Liu, H. Yang, M. Howard and C. Dean. **Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation.** *Proc Natl Acad Sci U S A.* 113(1): 218-23 (2016)

# Flowering time control: another window to the connection between antisense RNA and chromatin

Robert Ietswaart, Zhe Wu and Caroline Dean

Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK

A high proportion of all eukaryotic genes express antisense RNA (asRNA), which accumulates to varying degrees at different loci. Whether there is a general function for asRNA is unknown, but its widespread occurrence and frequent regulation by stress suggest an important role. The best-characterized plant gene exhibiting a complex antisense transcript pattern is the *Arabidopsis* floral regulator *FLOWERING LOCUS C* (*FLC*). Changes occur in the accumulation, splicing, and polyadenylation of this antisense transcript, termed *COOLAIR*, in different environments and genotypes. These changes are associated with altered chromatin regulation and differential *FLC* expression, provoking mechanistic comparisons with many well-studied loci in yeast and mammals. Detailed analysis of these specific examples may shed light on the complex interplay between asRNA and chromatin modifications in different genomes.

## Widespread antisense transcription

The transcriptome analysis of the *Arabidopsis thaliana* genome nearly a decade ago revealed, surprisingly, that approximately 30% of all the annotated genes exhibited significant asRNA expression [1]. This level of antisense transcript production has been a consistent observation in other whole-genome analyses, including *Drosophila*, human, and rice [2]. More recently, the application of next-generation sequencing methodologies has revealed the presence of pervasive transcription and long non-coding RNAs (lncRNAs, non-protein-coding RNA molecules at least 200 nt in length) in many genomes [3–6]. Transcription is clearly not limited to protein-coding genes and can be antisense to genes, divergent from promoters, convergent into genes, and completely intergenic [2]. The extent of lncRNAs in a genome seems to correlate with an increase in the complexity of the organism. For example, the protein-coding part constitutes only 1.5% of the human transcriptome, whereas in prokaryotes it accounts for over 75%. This has led to the suggestion that non-coding parts of a genome serve a gene regulatory role, which becomes increasingly important in higher organisms [7]. lncRNA can play important roles in gene regulation by acting as molecular signals, decoys, guides, and scaffolds [4]. It is

unclear at present whether *cis*-localized antisense transcripts, a subset of the more general long non-coding transcripts, play a functionally distinct role [8]. asRNAs have been proposed to function through transcriptional

## Glossary

**Autonomous pathway:** genetic pathway that controls flowering time in *Arabidopsis*. It is thought to contain several parallel activities all repressing *FLC*, independently of day length.

**Cis-NATs:** natural antisense transcripts transcribed from the same locus or a convergently transcribed locus.

**COLDAIR:** lncRNA expressed from within *FLC* intron 1 in the sense direction.

**COOLAIR:** lncRNA fully encompassing *FLC* in the antisense direction. It is alternatively polyadenylated and alternatively spliced.

**FCA:** RNA-binding protein in *Arabidopsis*, stimulating alternative polyadenylation genome-wide, and in particular of *COOLAIR*.

**FLD:** an *Arabidopsis* homolog of human histone demethylase LSD1. It represses *FLC* sense expression through reducing H3K4me2 levels in the body of the gene.

**FLO11:** a flocculin gene in yeast (~4 kb) with a variegated expression pattern due to the involvement of ICR1 and PWR1 sense and antisense lncRNAs, respectively.

**FLOWERING LOCUS C (FLC):** a well-studied *Arabidopsis* gene (~7 kb) that encodes a MADS-box transcription factor. This is a key repressor of the floral transition and is targeted by several regulatory pathways.

**FRIGIDA (FRI):** an *Arabidopsis* protein determining the need for vernalization through establishment of high *FLC* expression.

**FY:** *Arabidopsis* homolog of yeast Pfs2p and mammalian WDR33, components of the well-characterized pre-mRNA 3' end cleavage and polyadenylation stimulating factor complex (CPSF).

**Nonsense-mediated decay (NMD):** RNA decay pathway that targets aberrantly processed mRNAs. In plants it has recently been shown to target lncRNAs as well [22].

**Paf1C:** RNA polymerase II (RNAPII)-associated complex that resides at the C-terminal domain of RNAPII throughout transcription elongation. Conserved throughout eukaryotes, it has multiple functions in transcription, for example the recruitment of RNA processing factors and the facilitation of transcription elongation.

**Plant homeodomain (PHD):** a protein domain that is generally associated with the ability to recognize particular post-translationally modified histones.

**PHD-PRC2:** PRC2 accompanied by several PHD domain-containing proteins. It is required for establishing high H3K27me3 at *FLC* during vernalization.

**PHO84:** encodes a phosphate transporter gene in yeast (~2 kb), and is epigenetically silenced through the stabilization of asRNA during chronological aging.

**Polycomb repressive complex 2 (PRC2):** a widely conserved protein complex consisting of E(Z) (a histone methyltransferase), extra sex combs (ESC), p55, and Su(z)12. PRC2 catalyzes the methylation of histone tails at H3K27, leading to high levels of the repressive H3K27me3 mark.

**Tsix:** mammalian asRNA of *Xist* that regulates *Xist* expression in the process of X-chromosome inactivation.

**Vernalization:** a process in plants whereby flowering is accelerated by prolonged exposure to cold, ensuring that plants align their flowering with spring.

**Xist:** a mammalian non-coding RNA (~18 kb) that is key to female X-chromosome inactivation. Located on an X chromosome, it blankets that chromosome in a *cis*-limited fashion and is required for silencing.

**Xrn1:** conserved cytoplasmic 5'–3' exoribonuclease in yeast required for mRNA and ncRNA turnover.

Corresponding author: Dean, C. (caroline.dean@jic.ac.uk)

Keywords: antisense RNA; gene expression; chromatin regulation; *FLC*; non-coding RNA; *COOLAIR*.

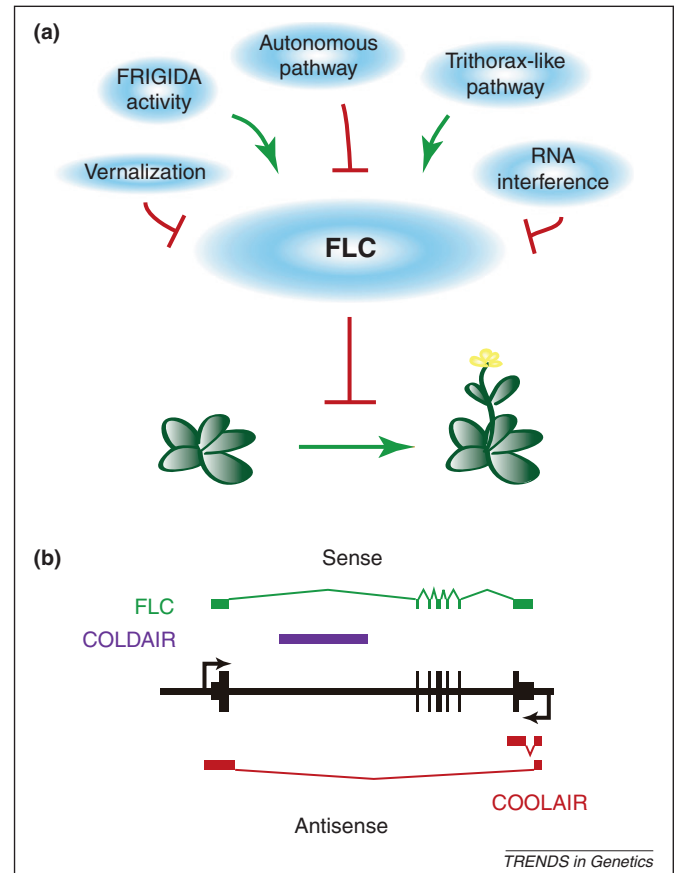
interference mechanisms [9], mediate chromatin modifications [10], or be involved in RNA interference mechanisms [11]. In *Saccharomyces cerevisiae* (yeast), they have been implicated in stress responsiveness and the regulation of inducible genes [12,13]. In addition, as a result of having lower baseline levels, genes expressing asRNA generally exhibit a wider expression range than those genes which do not express asRNA [14]. These characteristics and their abundance in a range of organisms suggest they could play a role in the plasticity of gene expression. In plants, this could be of profound importance for responses to environmental cues and adaptation. However, until we determine more mechanistic details for a range of examples from different organisms that will enable us to make inferences about generic mechanisms, their widespread role in genome regulation will remain an interesting possibility only.

In this review we introduce examples where the mechanism of antisense transcription has been analyzed and shown to play an important role in chromatin regulation. We focus on the best-characterized example in plants, the regulation of the floral repressor *FLC*, and then review specific examples in yeast and mammals to highlight conceptual parallels. Lastly, we speculate on the broader genomic roles of asRNAs and consider why some genes seem to have evolved into targets for asRNA regulation.

### *FLC* and antisense-mediated regulation in plants

The presence of extensive antisense transcription in several whole-genome transcriptome analyses was initially viewed with some skepticism and concern that the methodologies were subject to technical artifacts resulting from spurious second-strand complementary DNA synthesis in the reverse transcription reaction. This issue has been addressed directly in several studies [15,16] and extensive antisense transcription is still observed (although less so than in the early studies). The need for plants to constantly adapt to changing conditions makes a role for asRNA in gene regulation an attractive hypothesis [17], and it is clear from a genome-wide study that antisense transcripts in *Arabidopsis* show developmental, clock-related, or stress-related changes in expression [8,18]. The presence of asRNA has also been detected in rice, wheat, and legumes [19–21], but apart from an analysis of the role of the nonsense-mediated decay pathway (see Glossary) on their stability, relatively little mechanistic understanding is available [22]. Given the extensive understanding of small RNA (sRNA) pathways in plants, initial studies addressed whether natural antisense transcripts (NATs) transcribed from either the same locus or a convergently transcribed gene – so-called *cis*-NATs – triggered a double-stranded RNA mechanism involving sRNA production. However, *cis*-NATs do not show an increased sRNA signature compared to other genomic regions, and any correlation of presence of asRNA with low sense expression must therefore involve a different mechanism [23,24].

To date, the best-understood example of the role of antisense transcripts in plants is the regulation of the *Arabidopsis* floral repressor gene *FLC*. Here, the antisense expression independently intersects with two pathways that repress *FLC* expression (Figure 1). Below we detail our current understanding of this regulation.



**Figure 1.** The *FLC* gene. (a) *FLC* controls the transition from vegetative to reproductive development in *Arabidopsis* [25–28]. *FLC* expression is regulated by many pathways [61]: vernalization, a cold-induced epigenetic silencing that occurs during winter; FRIGIDA, a coiled-coil protein that upregulates *FLC* expression; the autonomous pathway, which is composed of many independent repressive activities; *Arabidopsis* Trithorax-like pathway, which stimulates *FLC* gene expression; RNA interference suppresses *FLC* in different *Arabidopsis* accessions [31,81]. (b) The *FLC* locus (~7 kb in length) expresses multiple types of transcripts. *FLC* mRNA encodes a MADS box transcription factor. *COOLAIR* is a non-coding transcript that fully encompasses *FLC* in the antisense direction. It is alternatively polyadenylated, with a proximal poly(A) site in sense intron 6 and a distal poly(A) site in the sense promoter region, and is differentially expressed in warm and cold conditions. *COLDAIR*, expressed under cold conditions from within intron 1 of *FLC* in the sense direction, is a capped but non-polyadenylated lncRNA.

*FLC* encodes a MADS box-type transcriptional repressor that prevents the activation of a suite of genes required for the floral transition [25–28]. The *FLC* antisense transcription unit fully encompasses the *FLC* gene (~7 kb in size), initiating from immediately downstream of the major sense poly(A) site, and terminating upstream of the sense initiation site (Figure 1). This led to the question of whether its initiation depends on the termination of the sense transcript. A fusion of the *COOLAIR* promoter to a luciferase coding region demonstrated that antisense transcription could be initiated and regulated independently of the sense transcript [29]. A similar independence of sense and antisense transcription has been found for yeast genes [30]. The RNA starts from several positions within a chromatin region marked by chromatin modification histone H3 lysine 9 dimethylation (H3K9me2) that is normally associated with heterochromatin in *Arabidopsis* and other organisms. This region also shows homology to sRNAs: a Dicer-dependent (24-mer) and a Dicer-independent 30-mer [31]. Analysis of the origin of these sRNAs led

to the identification of an antisense transcript that, owing to its upregulation by cold, was named *COOLAIR*, as a parallel to the lncRNA *HOTAIR* in *Drosophila* [29,32]. *COOLAIR* is now used to describe the *FLC* antisense transcript generally, both in warm- and cold-treated plants. *COOLAIR* transcription is generally positively correlated with sense transcription in a range of flowering mutants that both upregulate (late-flowering mutants) and downregulate (early-flowering mutants) *FLC* expression. However, this symmetry is broken during cold treatment – after 2 weeks of cold treatment, antisense levels increase 10-fold, whereas the *FLC* transcription is downregulated.

The *FLC* genetic regulatory network is complex, but a schematic illustration of the multiple activating and repressing pathways is shown in Figure 1. The activating pathways are conserved chromatin regulators including homologs of the Paf1 complex [33], RAD6-BRE1 [34–36], and the protein complex COMPASS [37–39] carrying *Arabidopsis* Trithorax homologs (*FLC* activating mechanisms are reviewed in [40]). SWR1, a conserved chromatin remodeler implicated in histone H2A.Z deposition, is also required for *FLC* upregulation [41]. In addition, the activator FRIGIDA contributes strongly to natural variation of *FLC* expression, and thus flowering. This is a coiled-coil, lysine-rich protein that interacts with a CAP-binding complex subunit *in vivo* and is required for the recruitment of chromatin modification machinery to the *FLC* locus [42–44]. These activators are antagonized by several repressors identified by genetic analysis of late-flowering mutants. These were grouped into the so-called autonomous pathway. However, instead of functioning in a linear genetic pathway, they constitute several parallel mechanisms that all repress *FLC*.

### Autonomous pathway

The autonomous pathway has been known to involve RNA regulation since *FCA* was shown to encode an RNA recognition motif (RRM) protein nearly 15 years ago [45]. *FCA* also has a WW protein interaction domain and this was used to identify the interacting protein *FY*, previously characterized through genetic analysis of a late-flowering mutant. *FY* is homologous to Pfs2p (yeast) and WDR33 (mammals), components of the well-characterized pre-mRNA 3'-end cleavage and polyadenylation stimulating factor (CPSF) [46–48]. *FCA* and *FY* were shown to be functionally important in RNA 3' processing as evidenced by their role in autoregulation of *FCA* polyadenylation site choice [49] and their genome-wide effect on polyadenylation [50]. However, mutations in *FCA* and *FY* do not influence *FLC* sense transcript 3' processing.

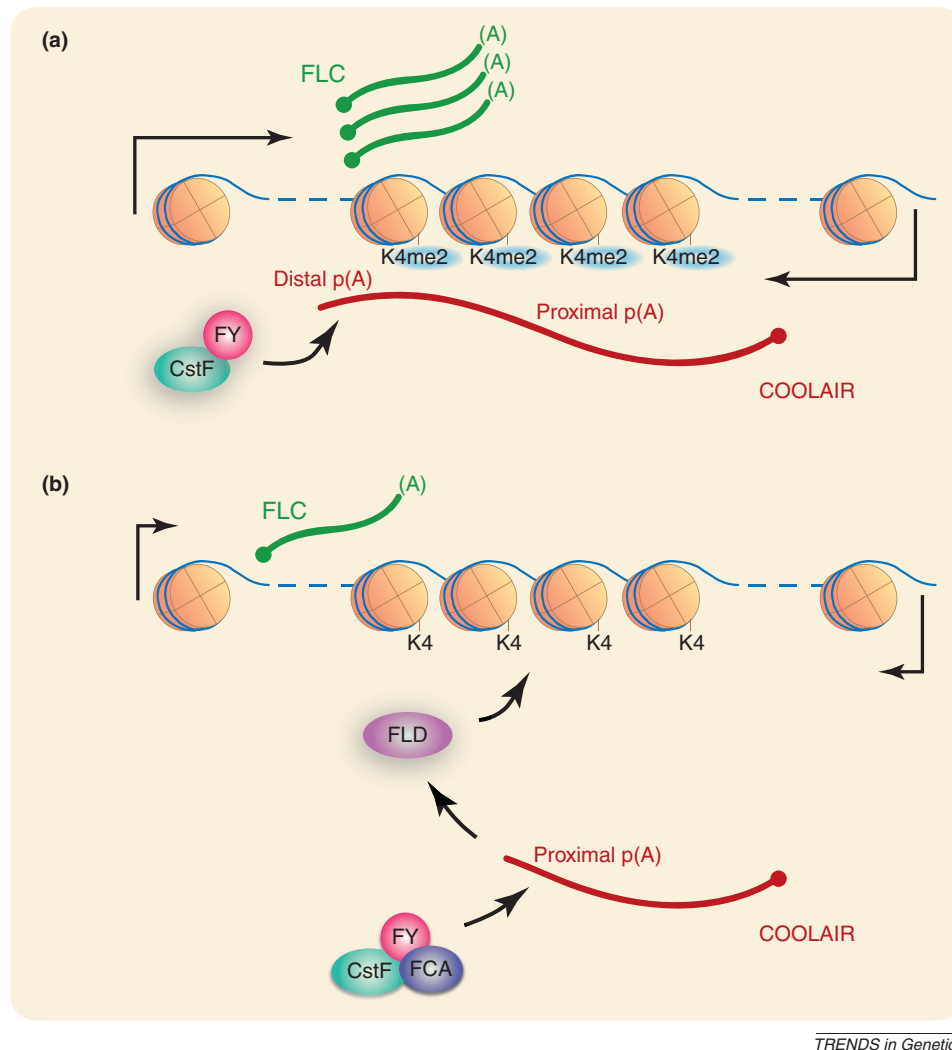
To understand how *FCA* activity results in a ~30-fold reduction in *FLC* expression, a suppressor mutagenesis was undertaken to identify all the required factors [51]. So far this has identified *FLD*, a homolog of the human lysine specific demethylase 1 (LSD1) [51,52], and two canonical 3'-processing factors, *CstF64* and *CstF77* [53]. Surprisingly, mutation of these conserved cotranscriptional machinery components, generally required to process RNA, did not reduce *FLC* production but instead increased it [53]. Analysis of 3' processing of *COOLAIR* provided insight into the mechanism; these mutations were affecting the

processing of the antisense transcript, which in turn increased *FLC* expression. *COOLAIR* is alternatively polyadenylated (Figure 1), as are many transcripts in the *Arabidopsis* genome [54], with both a proximal poly(A) site (within sense intron 6) and a distal poly(A) site (overlapping the sense promoter). *FCA* and *FY* function to promote the use of the proximal site, and *FCA* associates with *FLC* chromatin in this region [51]. Perhaps *FCA* interacts with *COOLAIR* through its RRM, bringing that transcript to the RNA 3'-processing machinery through the interaction with *FY* [55,56], thereby stimulating use of the proximal poly(A) site (Figure 2). *FPA*, another RRM protein that acts independently of *FY* [57], also promotes usage of the proximal site [58]. The choice of the proximal poly(A) site then triggers *FLD* demethylation to reduce the levels of dimethylated histone H3 lysine 4 (H3K4me2) in the body of *FLC*, leading to a repressed chromatin state and reduced sense and antisense expression [51,53]. How use of the proximal poly(A) site stimulates *FLD* activity remains to be elucidated. A relative increase in use of the distal poly(A) site in *fca*, *fpa*, or *fy* mutants (Figure 2) has led to the suggestion that increased antisense transcription through the sense promoter could stimulate sense transcription [58,59]. Opposing proximal poly(A) and distal poly(A) site choice could thus function antagonistically as an *FLC* repressor and activator, respectively. In this way regulated 3' processing of the *FLC* antisense transcript, *COOLAIR*, might modulate *FLC* expression. Further investigation is ongoing to dissect how changes in *COOLAIR* processing are linked to changes in transcription and chromatin structure. The conserved nature of several involved proteins suggests this could be a mechanism generally relevant for the many loci with antisense transcription.

### Vernalization

The autonomous pathway functions in parallel with vernalization, a second pathway repressing *FLC* expression [60]. Vernalization is a process whereby flowering is accelerated by prolonged cold, and it ensures that plants align their flowering with spring [61]. Prolonged cold represses *FLC* transcriptionally and induces epigenetic silencing that is mediated by a conserved Polycomb (PcG) mechanism [62,63]. *FLC* expression decreases in the cold in a quantitative manner, scaling with the length of cold the plants experience, and it remains epigenetically silenced during subsequent development after return to the warm [60,64]. Mutational analysis revealed that this process requires a modified PcG silencing complex including proteins containing a plant homeodomain (PHD). A core Polycomb repressive complex 2 (PRC2) associates with *FLC* chromatin independently of temperature [65]. After several weeks of cold, levels of histone H3 lysine 27 trimethylation (H3K27me3), which is associated with PcG silencing, accumulate at an intragenic site covering (sense) exon 1 and the 5' end of intron 1, referred to as the nucleation region (Figure 3). This correlates with accumulation of the modified PHD–PRC2 complex and its association with the nucleation site. H3K27me3 accumulates quantitatively at the nucleation region with increasing weeks in the cold [62,66]. Upon return to warm temperatures, the



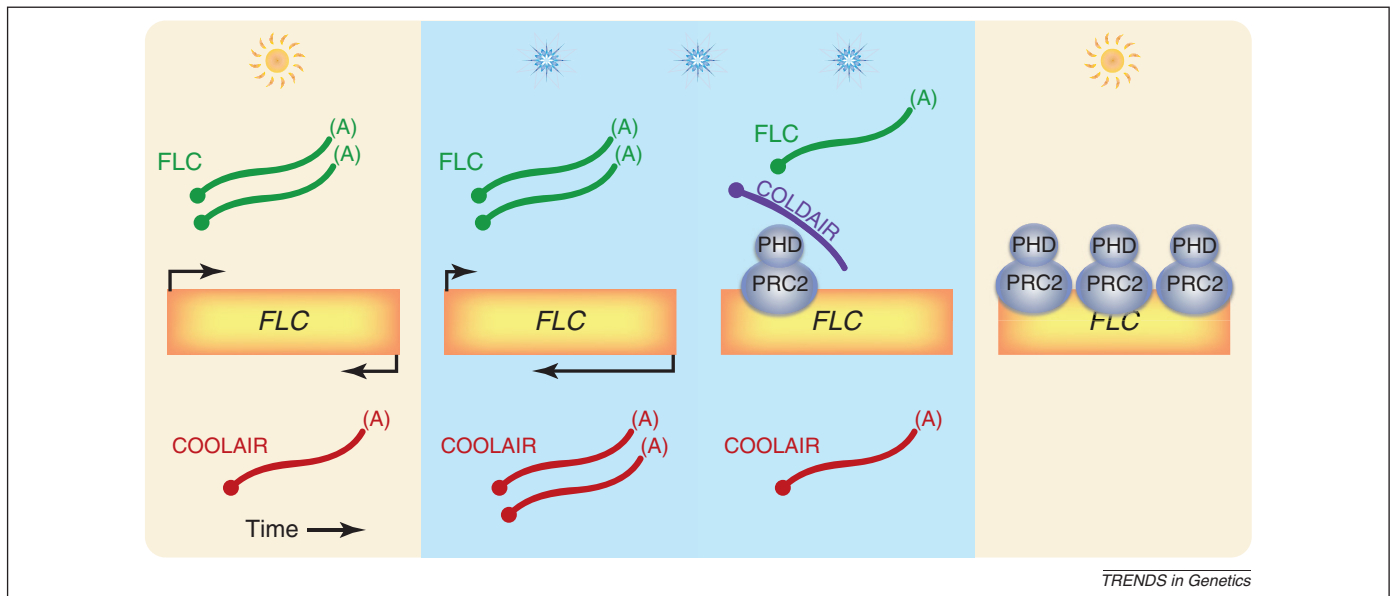


**Figure 2.** The autonomous pathway represses *FLC* expression through antisense RNA (asRNA)-mediated chromatin modifications. **(a)** In the absence of endogenous FCA, *COOLAIR* transcripts (red) are polyadenylated under the influence of FY, CstF64, and CstF77 at a distal poly(A) site. This is associated with high H3K4me2 (me2 in blue) levels in the body of the *FLC* gene, and with high levels of functional *FLC* mRNA (green) and *COOLAIR* expression. **(b)** FCA promotes alternative polyadenylation of *COOLAIR* through targeting FY and CstF activity to a proximal poly(A) site. This event is likely to trigger FLD-dependent demethylation in the body of *FLC*, leading to a transcriptionally repressed state and low levels of functional *FLC* mRNA [53].

PHD–PRC2 complex spreads across the whole *FLC* locus raising H3K27me3 to high levels, which are required to maintain the epigenetic stability of repression through subsequent cell divisions and environmental noise. This switch to the fully epigenetically silenced state is digital: it either occurs, and the gene is fully silenced; or it does not occur, and the gene is reactivated upon return to warm temperatures [67]. The probability of this cell-autonomous switch increases with length of cold such that the quantitative accumulation of epigenetic silencing by cold represents an increasing fraction of cells that have switched to the epigenetically silenced state [67]. How does *COOLAIR* function in this mechanism?

Using a custom microarray with single-nucleotide resolution of both strands of the *FLC* locus, changes in *FLC* transcripts at several phases of the vernalization process were identified [29]. The most striking difference was the accumulation of *FLC* antisense transcripts after plants had experienced cold (which peaked after 14 days). Unspliced and proximally polyadenylated *COOLAIR*

increased most significantly (~10-fold), but the distally polyadenylated form also increased slightly. The increase in antisense transcripts coincides with suppression of unspliced sense *FLC* levels, but not functional *FLC* mRNA levels, which take several more weeks of cold to decrease significantly. This led to the idea that *COOLAIR* may facilitate the decrease in *FLC* transcription during the first few weeks in the cold (Figure 3). Supporting this hypothesis, a transgene that contains the *COOLAIR* promoter fused downstream of a constitutively expressed green fluorescent protein (GFP) showed cold-induced antisense transcript production and cold-induced down-regulation of GFP expression [29]. Increased stability of the spliced *FLC* mRNA versus nascent transcripts might explain the observation that *FLC* mRNA levels remain unaffected during a two-week period of cold. When vernalized for longer, *COOLAIR* expression and *FLC* mRNA decrease over time and remain stably silenced after return to the warm [29,68]. In plants where T-DNA insertions attenuate production of *COOLAIR*, PcG silencing was not



**Figure 3.** Vernalization silences *FLC* epigenetically. Vernalization is the process whereby flowering is accelerated in response to the prolonged cold of winter. Time runs from left to right. Before exposure to cold, *FLC* (green) is expressed in higher amounts than *COOLAIR* (red). During the first 2 weeks of cold, *COOLAIR* is upregulated whereas unspliced *FLC* RNA, commonly thought to reflect sense transcription, decreases [29]. However, the functional *FLC* mRNA levels are still similar to non-vernalized conditions, possibly due to differences in lifetimes between nascent and processed RNAs. During prolonged cold, *FLC* is progressively epigenetically silenced via a Polycomb-dependent mechanism correlating with *COLDIAIR* expression [62,67,68]. Functional PHD–PRC2 complexes associate with the nucleation region slightly downstream of the sense promoter leading to an upregulation in the repressive histone mark H3K27me3 at this site. After returning to warm conditions, the PHD–PRC2 complex associates across the whole locus, leading to high levels of H3K27me3 across the gene. The high H3K27me3 levels are required for the epigenetic stability of *FLC* throughout the rest of development.

disrupted after 4 weeks of cold treatment, showing that antisense transcripts are not required for vernalization [69]. However, this does not preclude a possible function for *COOLAIR* during the first 2 weeks of vernalization preceding epigenetic silencing [29]. Indeed, antisense transcripts seem to influence the dynamics of transcriptional repression in this early cold period (Q. Sun and C. Dean, unpublished). This might be explained by the model discussed in the previous section: a relative increase in a repressive proximal antisense transcript compared to an activating distal transcript would result in the overall repression of sense *FLC*.

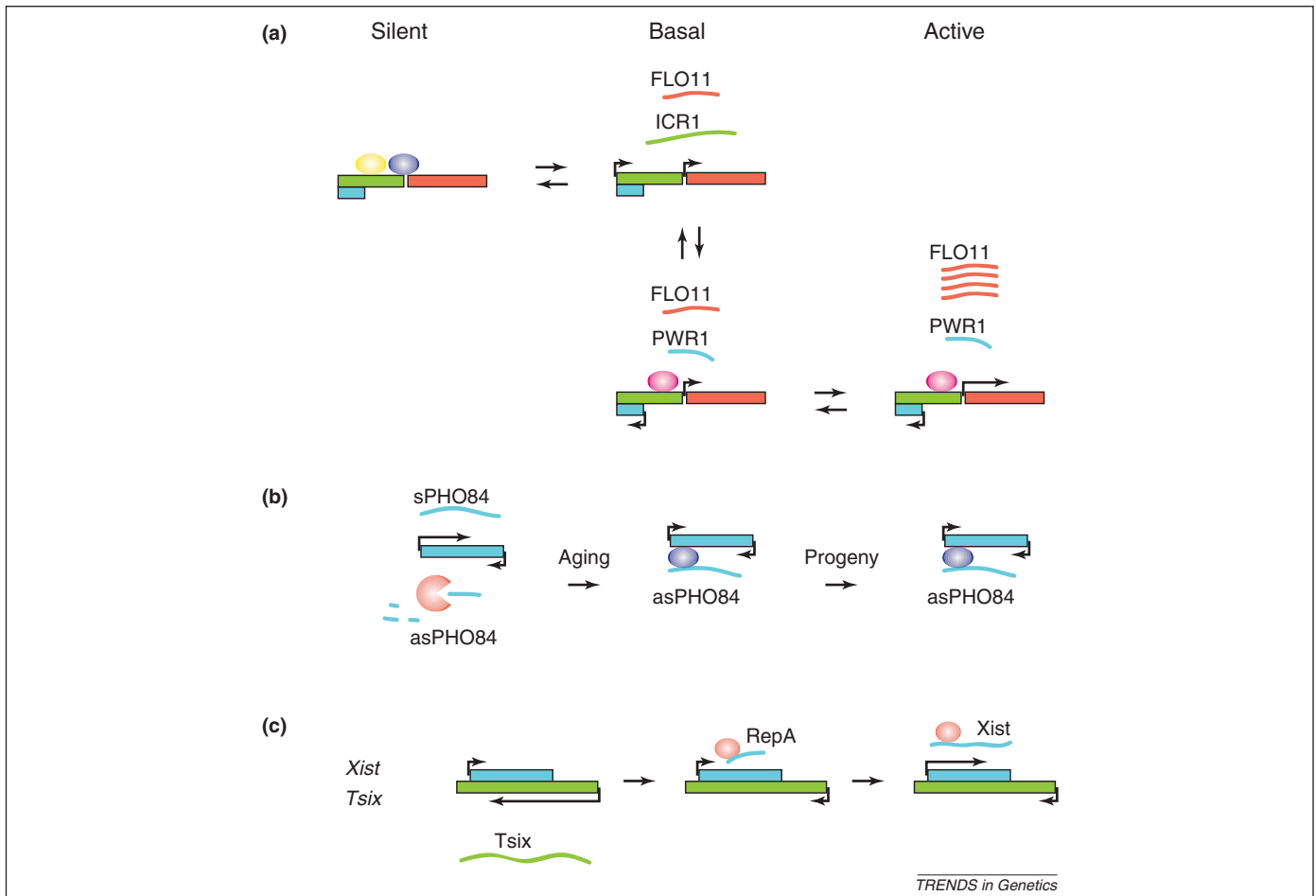
A sense lncRNA (~1 kb), termed *COLDIAIR*, has been detected within *FLC* intron 1 (Figure 1). It contains a 5' cap but is not polyadenylated and is induced by cold, but later than *COOLAIR* – reaching maximum levels after 3 weeks of cold. *COLDIAIR* was found to be associated with the core PRC2 components, and its knockdown led to attenuation of epigenetic silencing, suggesting that it plays an important role in triggering PcG silencing [68]. In summary, there appears to be a complex interplay between antisense transcripts, other lncRNAs, and chromatin-modifying complexes in the various steps of the vernalization process (Figure 3). It will take integration of genetic, biochemical, and modeling approaches to unravel the mechanistically redundant functions contributing to this mechanism.

The understanding of antisense-mediated gene regulation is much further advanced in yeast and mammals. We discuss some well-studied examples below, which provide important concepts for the broader roles of asRNA in chromatin regulation and, to gain further insight into this connection, we suggest some possible mechanistic links between *FLC* and these systems.

### Genetic toggle mechanism

A universal issue in asRNA-chromatin regulation is to understand whether the RNA molecule itself or the act of transcription is important for the regulation [70]. The dissection of the regulation of the yeast *FLO11* gene suggests that for some examples it is the latter. *FLO11* encodes a glycoprotein important for adhesion, and it is regulated by a pair of antagonistic lncRNAs – resulting in a variegated gene expression pattern that is likely to give considerable selective advantage to cells as the environment changes [71–73]. A sense lncRNA (ICR1) is transcribed upstream of the *FLO11* promoter. Antisense to this is another lncRNA, PWR1 (Figure 4). Competitive binding of the activator (Flo8) or repressor (Sfl1) to the *FLO11* promoter determines which of the two lncRNAs is transcribed. Binding of Flo8 stimulates PWR1 transcription and inhibits ICR1 expression, establishing a basal expression state of *FLO11*, which can be upregulated by conventional transactivating factors. ICR1 transcription is thought to clear the *FLO11* promoter, enabling binding of Sfl1, which then recruits a histone deacetylation complex (HDAC) to silence the gene fully. The overall effect is variegated transcription in individual cells, monitored in a recent study using single-molecule RNA fluorescence *in situ* hybridization (FISH) [71,73], with three expression states: silenced, basal, or active expression. The promoter-localized lncRNAs – one of course antisense to the other – thus provide a genetic toggle that contributes to the observed variegated expression of *FLO11*. This concept is appealing in the *FLC* context: instead of a toggle mechanism determining which promoter lncRNA is expressed as in the case of *FLO11*, the choice of either the proximal or distal antisense poly(A) site could be a toggle mechanism





**Figure 4.** Antisense RNA can regulate gene expression in yeast and mammals. (a) *FLO11* (red gene) expression in yeast can be in three states: silent, basal, or active. The silent state is associated with binding of repressor Sfl1 (yellow protein), which recruits an HDAC (purple protein) to the promoter, leading to transcriptional silencing of *FLO11*, the sense ncRNA *ICR1* (green gene), and asRNA *PWR1* (antisense direction, blue gene). In the basal state, expression of the ncRNAs toggles between *ICR1* and *PWR1*, which is associated with basal low expression of *FLO11*. *ICR1* overlaps with the *FLO11* promoter, and its transcription is thought to clear the promoter of binding proteins, allowing competitive binding between Sfl1 and the activator Flo8 (pink protein). Conversely, *PWR1* expression interferes with *ICR1* and is associated with *FLO11* activation because it can coalesce with Flo8 binding and recruit more transcriptional machinery. This results in an active state with high *FLO11* levels [71,73]. (b) Another yeast gene, *PHO84* (blue gene), is epigenetically silenced in the process of chronological aging. Sense expression (*sPHO84*) is initially high. As a result of reduced nuclear exosome activity over time, higher antisense *PHO84* (*asPHO84*) expression induces recruitment of HDACs (purple protein) to the sense promoter. This epigenetically silenced state is inherited by the progeny [10]. (c) Before the initiation of female X-chromosome inactivation (XCI), the *Xist* (blue gene) is silenced because *Tsix* (green gene) expression induces DNA methylation at the *Xist* promoter. At the initiation of XCI on the chromosome that is to be inactivated, *Tsix* is downregulated, and it is proposed that RepA, an lncRNA located within *Xist* in the sense direction, targets PRC2 complexes (red protein) to the locus. At the inactivated X chromosome, *Xist* is expressed and associates with PRC2 complexes and spreads in a *cis*-limited fashion across the inactivated X chromosome [77,78].

for *FLC*. This switch may then be the signal for recruitment of chromatin machinery necessary for low and high *FLC* expression states.

#### RNA turnover linked to histone modifications

Another classic example of a functionally important antisense transcript in yeast is that of the *PHO84* gene, which encodes a phosphate transporter. *PHO84* asRNA originates from a convergent promoter at the 3' end of the gene (Figure 4). Chronological aging of yeast cells, a process influenced by cold, was found to induce this asRNA, and this resulted in silencing of *PHO84* sense transcription in an epigenetically stable manner [10]. This arose from inactivation of the nuclear exosome over time, which enabled the asRNA transcripts to accumulate and mediate HDAC recruitment to the *PHO84* promoter, leading to epigenetic silencing. These data suggest that the RNA itself is important for this mechanism. Introduction of an ectopic *PHO84* gene copy led to silencing of both the

endogenous and ectopic gene mediated by its asRNA. However, this was via a somewhat different mechanism, not requiring HDACs but involving the H3K4 methyltransferase Set1, that stimulates antisense transcription [74]. In yeast, chronological aging reduces exosome activity, stabilizing antisense transcripts. This raises the question: could environmental factors alter exosome function thereby triggering chromatin modifications and gene expression changes? It will be interesting to determine if cold during the first few weeks of vernalization functions through such a mechanism to upregulate *COOLAIR*.

Different RNA turnover mechanisms appear to delineate functional ncRNA groups in yeast [75]. One class of over 1000 ncRNAs antisense to yeast open reading frames (ORFs) are degraded by the cytoplasmic 5'–3' exoribonuclease Xrn1 [12]. These Xrn1-sensitive unstable transcripts (XUTs) are polyadenylated and transcribed by RNA polymerase II, for example at the *TIR1* locus and retrotransposon *TY1* [76]. In *xrn1* mutants, *TIR1*axut and

RTL (the antisense transcripts of *TIR1* and *TY1*, respectively) are stabilized, and this correlates with decreased sense gene expression. RTL also mediates chromatin silencing in *trans*, partially through histone deacetylation and histone methylation [76], with H3K4me and H3K4me2 being important for XUT-mediated silencing [12]. H3K4me3 antagonizes XUT activity, and this may result in the antisense transcripts mediating gene silencing in a chromatin-dependent manner: in the presence of elevated H3K4me3 levels at the sense promoter, a hallmark of active chromatin, stabilization of XUTs seems to have less of an effect on sense gene expression [12]. This reinforces the suggestion that asRNA-mediated repression predominates when sense expression is already low, leading to an increased expression range due to lower baseline levels in the repressed state [14].

### Antisense RNA in Polycomb silencing

There has also been a large amount of analysis on the role of specific sense/antisense lncRNAs in chromatin regulation in mammals, the most well-known pair being *Xist* and *Tsix* that play a central role in the choice of which X chromosome is inactivated in female mammals [77,78]. *Xist* encodes an 18 kb nuclear RNA essential for X-chromosome inactivation (XCI). Before the initiation of XCI, *Xist* is antagonized by biallelic transcription of *Tsix*, its antisense counterpart (Figure 4). This silencing is not due to occlusion of the transcription machinery at the *Xist* promoter by *Tsix* transcription (as suggested in the genetic toggle mechanism described above) but instead by *Tsix* RNA associating with DNA methylation machinery to stably silence the *Xist* promoter. It is proposed that, during X-chromosome pairing, downregulation of *Tsix* on the future inactive X chromosome allows *RepA*, a sense lncRNA located within an *Xist* exon, to target PRC2 to *Xist*. *Xist*-PRC2 complexes tethered to the chromatin then spread in a *cis*-limited fashion to cover the whole X chromosome, resulting in its inactivation. Similar principles of transcription from both strands, chromosome coating, and association with chromatin machinery, have also been demonstrated for lncRNAs involved in genomic imprinting in mammals [70,79,80]. These RNAs silence a set of genes in an allele-specific manner such that these genes inherited through the maternal and paternal gametes are differentially expressed. A well-characterized example is *Kcnq1ot1*, which encodes a 91 kb transcript antisense to the *Kcnq1* gene cluster. This RNA coats a ~1 Mb region of the paternal chromosome, resulting in the epigenetic silencing of 8–10 genes in this *Kcnq1* domain that remain expressed from the maternal alleles.

From the discussion above, describing *Xist*, *Tsix*, *RepA*, and *Kcnq1ot1*, as well as the involvement of *COOLAIR* and *COLDAR* in regulating *FLC* during vernalization, it is clear that sense and antisense lncRNAs are involved in PcG-mediated silencing of genes in both plants and mammals. Although the details of each regulatory system appear different at this stage, additional conceptual parallels are likely to emerge from further investigation.

### Concluding remarks

The emerging view is that many genes will show some non-coding transcription. A key question is now whether

different kinds of ncRNAs represent different functional groups, defined by long versus short, antisense versus sense, or through their type of degradation pathway. We have focused in this review on the function of antisense lncRNAs in plant gene regulation with comparison to well-characterized examples in yeast and mammals. The involvement of asRNA in multiple pathways regulating *FLC* raises the question of why some genes become targets for asRNA regulation. Could these be genes where expression heterogeneity provided a very strong selective advantage, or where very low expression levels are required? asRNAs seem to play a major role in generating expression heterogeneity at their corresponding loci. The genetic toggle mechanism, elaborated for *FLO11* [71], is an attractive model to account for this. The quantitative variation in expression that antisense regulation confers could be very important in many natural contexts. With respect to flowering-time control, subtle changes in timing have large implications for seed yield and thus reproductive success. We might see many environmental inputs influencing the antisense-mediated regulation of *FLC* in specific genotypes, and also find extensive variation in this mechanism in natural populations adapted to very different climates.

The evidence for pervasive transcription suggests that transcriptionally silenced states are not the default, raising the need for active repression mechanisms. asRNAs may enhance repressive chromatin to reduce low sense transcription to a fully repressed state [14]. Dissection of the complexity of antisense-mediated regulation of *FLC* and other important plant targets, aided by concepts from many other systems, will be key areas for future investigations.

### Acknowledgments

The authors apologize to those whose work on the relation between asRNA and chromatin could not be acknowledged owing to space constraints. This work is supported by funding from Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/G01406X/1 (Z.W.), VSBfonds Scholarship and Prins Bernhard Cultuurfonds Scholarship to R.I., and European Commission AENEAS funding to C.D. We thank members of the Dean laboratory for critical reading of the manuscript.

### References

- 1 Yamada, K. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302, 842–846
- 2 Kapranov, P. *et al.* (2007) Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 8, 413–423
- 3 Wilusz, J.E. *et al.* (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23, 1494–1504
- 4 Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914
- 5 Ponting, C.P. *et al.* (2009) Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641
- 6 Eggleston, A.K. *et al.* (2012) Regulatory RNA. *Nature* 482, 321
- 7 Mattick, J.S. (2004) RNA regulation: a new genetics? *Nat. Rev. Genet.* 5, 316–323
- 8 Okamoto, M. and Seki, M. (2011) Expression profile and 5'-terminal structure of *Arabidopsis* antisense transcripts expressed in seeds. *Plant Signal. Behav.* 6, 691–693
- 9 Shearwin, K.E. *et al.* (2005) Transcriptional interference – a crash course. *Trends Genet.* 21, 339–345
- 10 Camblong, J. *et al.* (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131, 706–717

- 11 Matzke, M.A. and Birchler, J.A. (2005) RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.* 6, 24–35
- 12 van Dijk, E.L. *et al.* (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114–117
- 13 Geisler, S. *et al.* (2012) Decapping of long noncoding RNAs regulates inducible genes. *Mol. Cell* 45, 1–13
- 14 Xu, Z. *et al.* (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.* 7, 468
- 15 Poole, R.L. *et al.* (2008) Analysis of wheat SAGE tags reveals evidence for widespread antisense transcription. *BMC Genomics* 9, 475
- 16 Peroochi, F. *et al.* (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* 35, e128
- 17 Kuhn, J.M. *et al.* (2007) mRNA metabolism of flowering-time regulators in wild-type *Arabidopsis* revealed by a nuclear cap binding protein mutant, *abh1*. *Plant J.* 50, 1049–1062
- 18 Hazen, S.P. *et al.* (2009) Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol.* 10, R17
- 19 Zhou, X.F. *et al.* (2009) Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. *Genome Res.* 19, 70–78
- 20 Coram, T.E. *et al.* (2009) Large-scale analysis of antisense transcription in wheat using the Affymetrix GeneChip Wheat Genome Array. *BMC Genomics* 10, 253
- 21 Bardou, F. *et al.* (2011) Dual RNAs in plants. *Biochimie* 93, 1950–1954
- 22 Kurihara, Y. *et al.* (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2453–2458
- 23 Jen, C.H. *et al.* (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol.* 6, R51
- 24 Henz, S.R. *et al.* (2007) Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* 144, 1247–1255
- 25 Michaels, S.D. and Amasino, R.M. (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–956
- 26 Bäurle, I. and Dean, C. (2006) The timing of developmental transitions in plants. *Cell* 125, 655–664
- 27 Fornara, F. *et al.* (2010) SnapShot: control of flowering in *Arabidopsis*. *Cell* 141 550–550.e2
- 28 Deng, W. *et al.* (2011) *FLOWERING LOCUS C* (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6680–6685
- 29 Swiezewski, S. *et al.* (2009) Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 462, 799–802
- 30 Murray, S.C. *et al.* (2011) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res.* 40, 2432–2444
- 31 Swiezewski, S. *et al.* (2007) Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, *FLC*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3633–3638
- 32 Rinn, J.L. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323
- 33 Oh, S. *et al.* (2004) A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the *Arabidopsis FLC/MAF* MADS box gene family. *Plant Cell* 16, 2940–2953
- 34 Cao, Y. *et al.* (2008) Histone H2B monoubiquitination in the chromatin of *FLOWERING LOCUS C* regulates flowering time in *Arabidopsis*. *Plant Cell* 20, 2586–2602
- 35 Xu, L. *et al.* (2009) The E2 ubiquitin-conjugating enzymes, AtUBC1 and AtUBC2, play redundant roles and are involved in activation of FLC expression and repression of flowering in *Arabidopsis thaliana*. *Plant J.* 57, 279–288
- 36 Gu, X. *et al.* (2009) Repression of the floral transition via histone H2B monoubiquitination. *Plant J.* 57, 522–533
- 37 Pien, S. *et al.* (2008) *ARABIDOPSIS TRITHORAX1* dynamically regulates *FLOWERING LOCUS C* activation via histone 3 lysine 4 trimethylation. *Plant Cell* 20, 580–588
- 38 Tamada, Y. *et al.* (2009) *ARABIDOPSIS TRITHORAX-RELATED7* is required for methylation of lysine 4 of histone H3 and for transcriptional activation of *FLOWERING LOCUS C*. *Plant Cell* 21, 3257–3269
- 39 Jiang, D.H. *et al.* (2011) *Arabidopsis* COMPASS-Like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet.* 7, e1001330
- 40 Crevillén, P. and Dean, C. (2011) Regulation of the floral repressor gene *FLC*: the complexity of transcription in a chromatin context. *Curr. Opin. Plant Biol.* 14, 38–44
- 41 Deal, R.B. *et al.* (2005) The nuclear actin-related protein ARP6 is a pleiotropic developmental regulator required for the maintenance of *FLOWERING LOCUS C* expression and repression of flowering in *Arabidopsis*. *Plant Cell* 17, 2633–2646
- 42 Johanson, U. *et al.* (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290, 344–347
- 43 Choi, K. *et al.* (2011) The *FRIGIDA* complex activates transcription of *FLC*, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. *Plant Cell* 23, 289–303
- 44 Ko, J.H. *et al.* (2010) Growth habit determination by the balance of histone methylation activities in *Arabidopsis*. *EMBO J.* 29, 3208–3215
- 45 Macknight, R. *et al.* (1997) *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* 89, 737–745
- 46 Simpson, G.G. *et al.* (2003) FY is an RNA 3' end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* 113, 777–787
- 47 Ohnacker, M. *et al.* (2000) The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. *EMBO J.* 19, 37–47
- 48 Shi, Y.S. *et al.* (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* 33, 365–376
- 49 Quesada, V. *et al.* (2003) Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J.* 22, 3142–3152
- 50 Sonmez, C. *et al.* (2011) RNA 3' processing functions of *Arabidopsis* FCA and FPA limit intergenic transcription. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8508–8513
- 51 Liu, F. *et al.* (2007) The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. *Mol. Cell* 28, 398–407
- 52 Shi, Y.J. *et al.* (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953
- 53 Liu, F. *et al.* (2010) Targeted 3' processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. *Science* 327, 94–97
- 54 Wu, X. *et al.* (2011) Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12533–12538
- 55 Henderson, I.R. *et al.* (2005) An allelic series reveals essential roles for FY in plant development in addition to flowering-time control. *Development* 132, 3597–3607
- 56 Manzano, D. *et al.* (2009) Altered interactions within FY/AtCPSF complexes required for *Arabidopsis* FCA-mediated chromatin silencing. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8772–8777
- 57 Bäurle, I. and Dean, C. (2008) Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of *Arabidopsis* targets. *PLoS ONE* 3, e2733
- 58 Hornyik, C. *et al.* (2010) The Spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev. Cell* 18, 203–213
- 59 Rosonina, E. and Manley, J.L. (2010) Alternative polyadenylation blooms. *Dev. Cell* 18, 172–174
- 60 Sheldon, C.C. *et al.* (2000) The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). *Proc. Natl. Acad. Sci. U.S.A.* 97, 3753–3758
- 61 Kim, D.-H. *et al.* (2009) Vernalization: winter and the timing of flowering in plants. *Annu. Rev. Cell Dev. Biol.* 25, 277–299
- 62 De Lucia, F. *et al.* (2008) A PHD–polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16831–16836
- 63 Buzas, D.M. *et al.* (2011) Transcription-dependence of histone H3 lysine 27 trimethylation at the *Arabidopsis* polycomb target gene *FLC*. *Plant J.* 65, 872–881

- 64 Sheldon, C.C. *et al.* (2002) Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. *Plant Cell* 14, 2527–2537
- 65 Wood, C.C. *et al.* (2006) The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14631–14636
- 66 Finnegan, E.J. and Dennis, E.S. (2007) Vernalization-induced trimethylation of histone H3 lysine 27 at *FLC* is not maintained in mitotically quiescent cells. *Curr. Biol.* 17, 1978–1983
- 67 Angel, A. *et al.* (2011) A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* 476, 105–108
- 68 Heo, J.B. and Sung, S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76–79
- 69 Helliwell, C.a. *et al.* (2011) Vernalization-repression of *Arabidopsis* *FLC* requires promoter sequences but not antisense transcripts. *PLoS ONE* 6, e21513
- 70 Mohammad, F. *et al.* (2009) Epigenetics of imprinted long noncoding RNAs. *Epigenetics* 4, 277–286
- 71 Tuck, A.C. and Tollervey, D. (2012) An RNA reset button. *Mol. Cell* 45, 435–436
- 72 Bumgarner, S.L. *et al.* (2009) Toggle involving *cis*-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18321–18326
- 73 Bumgarner, S.L. *et al.* (2012) Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol. Cell* 45, 470–482
- 74 Camblong, J. *et al.* (2009) *Trans*-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev.* 23, 1534–1545
- 75 Tisseur, M. *et al.* (2011) Pervasive transcription – lessons from yeast. *Biochimie* 93, 1889–1896
- 76 Berretta, J. *et al.* (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev.* 22, 615–626
- 77 Lee, J.T. (2011) Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat. Rev. Mol. Cell Biol.* 12, 815–826
- 78 Augui, S. *et al.* (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* 12, 429–442
- 79 Redrup, L. *et al.* (2009) The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* 136, 525–530
- 80 Pandey, R.R. *et al.* (2008) *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level Regulation. *Mol. Cell* 32, 232–246
- 81 Liu, J. *et al.* (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* 18, 2873–2878



# Competing ParA Structures Space Bacterial Plasmids Equally over the Nucleoid

Robert Ietswaart<sup>1,9</sup>, Florian Szardenings<sup>2,9</sup>, Kenn Gerdes<sup>2,3</sup>, Martin Howard<sup>1\*</sup>

**1** Computational and Systems Biology, John Innes Centre, Norwich, United Kingdom, **2** Centre for Bacterial Cell Biology, Newcastle University, Newcastle upon Tyne, United Kingdom, **3** Department of Biology, University of Copenhagen, Copenhagen, Denmark



## Abstract

Low copy number plasmids in bacteria require segregation for stable inheritance through cell division. This is often achieved by a *parABC* locus, comprising an ATPase ParA, DNA-binding protein ParB and a *parC* region, encoding ParB-binding sites. These minimal components space plasmids equally over the nucleoid, yet the underlying mechanism is not understood. Here we investigate a model where ParA-ATP can dynamically associate to the nucleoid and is hydrolyzed by plasmid-associated ParB, thereby creating nucleoid-bound, self-organizing ParA concentration gradients. We show mathematically that differences between competing ParA concentrations on either side of a plasmid can specify regular plasmid positioning. Such positioning can be achieved regardless of the exact mechanism of plasmid movement, including plasmid diffusion with ParA-mediated immobilization or directed plasmid motion induced by ParB/*parC*-stimulated ParA structure disassembly. However, we find experimentally that *parABC* from *Escherichia coli* plasmid pB171 increases plasmid mobility, inconsistent with diffusion/immobilization. Instead our observations favor directed plasmid motion. Our model predicts less oscillatory ParA dynamics than previously believed, a prediction we verify experimentally. We also show that ParA localization and plasmid positioning depend on the underlying nucleoid morphology, indicating that the chromosomal architecture constrains ParA structure formation. Our directed motion model unifies previously contradictory models for plasmid segregation and provides a robust mechanistic basis for self-organized plasmid spacing that may be widely applicable.

**Citation:** Ietswaart R, Szardenings F, Gerdes K, Howard M (2014) Competing ParA Structures Space Bacterial Plasmids Equally over the Nucleoid. PLoS Comput Biol 10(12): e1004009. doi:10.1371/journal.pcbi.1004009

**Editor:** Christopher V. Rao, University of Illinois at Urbana-Champaign, United States of America

**Received:** June 30, 2014; **Accepted:** October 28, 2014; **Published:** December 18, 2014

**Copyright:** © 2014 Ietswaart et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by a Biotechnology and Biological Sciences Research Council studentship, VSBfonds Scholarship and Prins Bernhard Cultuurfonds Scholarship (RI) and by grant BB/J004561/1 from the BBSRC (MH) and grant 089733/Z/09/Z from the Wellcome Trust (KG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: martin.howard@jic.ac.uk

<sup>9</sup> These authors contributed equally to this work.

## Introduction

*parABC* loci generate equally spaced positioning of many bacterial low copy number plasmids, thereby ensuring stable plasmid inheritance [1]. However, the underlying mechanism of action is not satisfactorily understood. In contrast, plasmid segregation mediated by actin homolog ParM is increasingly well explained and involves filaments that push plasmids apart in a mitotic-like process [2]. Understanding of the *parABC* mechanism is important, as it belongs to the most common class of DNA segregation systems in prokaryotes, used by chromosomes and antibiotic-resistance-carrying plasmids [1,3–5]. Moreover, it is used in other conceptually similar processes, such as chemotactic cluster positioning and partitioning of carbon-fixing carboxysomes [6,7].

The *parABC* locus present in *Escherichia coli* plasmids such as pB171 and P1 encodes two proteins: ParA, a P-loop ATPase that binds DNA non-specifically in its dimeric ATP-bound form (ParA-ATP for short) [8,9], and the DNA-binding protein ParB that binds site-specifically to the *parC* region [10,11]. Fluorescence microscopy has provided evidence for ParA movement over the

nucleoid with spatiotemporal oscillations in helix-like structures [12–14]. ParB and *parC* are required for these dynamics [12], with ParB promoting the conversion of ParA-ATP to dimeric ParA-ADP (ParA-ADP for short), causing ParA to unbind from the nucleoid [8,9]. The time period required for nucleoid-disassociated ParA to regain the ability to bind the nucleoid is sufficiently long *in vitro* to ensure that the relative locations of ParA-ADP unbinding and later ParA-ATP rebinding would be uncorrelated due to cytoplasmic ParA diffusion [8]. However, once nucleoid-bound, whether ParA-ATP then polymerizes to form long filaments *in vivo* is currently controversial. Furthermore, the means by which plasmids move under the influence of ParA, and whether ParA polymerization is important for this movement, are also unclear. Nevertheless, the outcome of these ParA dynamics in *E. coli* is equally spaced positioning of plasmid foci over the nucleoid [9,13–15]. This state is achieved regardless of the plasmid focus number  $n_p$  or cell length, with plasmid foci repositioned in the wake of retracting ParA structures [9].

Several mechanisms have been proposed to explain ParA-mediated plasmid movement. One hypothesis proposes that ParA-ATP polymerizes on the nucleoid to form long filaments and that

## Author Summary

How DNA is stably inherited through cell division is a fundamental question in cell biology. The most common system that mediates plasmid DNA inheritance in bacteria is through a *parABC* locus, encoding proteins ParA and ParB, and DNA sequence *parC*. These components can position plasmids at equally spaced positions throughout a cell to ensure plasmids are present in both daughter cells when the cell divides into two. Here we study the mechanism by which ParA structures achieve this precise positioning. We show that ParA can direct relatively immobile plasmids over the bacterial chromosome using self-organizing, competitive ParA structures, whose disassembly is induced by plasmid *parC*-bound ParB. More generally these findings will help us to understand transport and regular positioning of intracellular cargo.

plasmid translocation is achieved by ParB-stimulated retraction of the polymers, generating effective plasmid-pulling [3,9]. Other proposals are based on ParA-ATP forming a gradient-like distribution on the nucleoid, without a necessity for polymerization [8,16–21]. It is currently unclear whether any of these mechanisms can explain equal plasmid spacing given the known physiological and biochemical constraints. Here, we therefore investigate which aspects of the polymer and gradient mechanisms are required and sufficient to explain the observed plasmid translocation and equal spacing over the nucleoid.

We begin by showing mathematically that competition between dynamic ParA concentrations on either side of a plasmid can lead to equal plasmid spacing. This mechanism relies on an ability of a plasmid to move towards higher ParA concentrations, but the exact means of such movement is not important. We then investigate theoretically specific means of plasmid movement and examine whether predictions from such models are borne out experimentally. We define a computational diffusion/immobilization model where nucleoid-bound ParA-ATP can anchor diffusing plasmids. We show that diffusion/immobilization can in principle space mobile plasmids equally over the nucleoid. However, experiments measuring increased plasmid mobility in the presence of the pB171 *parABC* locus (*par2*), lead us to disfavor this model. Instead we favour a directed motion mechanism in which ParA structure formation provides directionality to plasmid motion thereby speeding up plasmid movement. The directed motion model produces robust equal plasmid spacing with, on average, relatively symmetric ParA distributions, a prediction we verify experimentally. Furthermore, we show experimentally that ParA organization is dependent on the underlying nucleoid structure, with nucleoid disruption resulting in perturbed plasmid positioning. Our combination of modeling and experiments has for the first time uncovered a robust mechanism for plasmid spacing that unifies previous proposals.

## Results

### ParB-GFP foci are spaced equally over the nucleoid

To study *par2*-mediated plasmid segregation, we investigated ParB-GFP localization, expressed from a *par2*-carrying mini-R1 test plasmid. The *par2* locus containing the *parB::sfGFP* fusion is fully functional as judged by loss-frequency assays (S1A Fig.). As previously described, usage of ParA-GFP and the *tetO*-TetR-mCherry labeling system also does not affect plasmid stability, indicating full functionality [9,12]. ParB-GFP forms foci that are

regularly positioned along the long cell axis *in vivo* (Fig. 1A), consistent with ParB-binding to plasmid-encoded *parC* regions [10,11]. Since plasmid dynamics occur primarily over the nucleoid, we reasoned that plasmid positioning with respect to the nucleoid rather than cell length is most informative. Therefore we measured ParB-GFP foci localization, together with Hoechst (DNA) stain to determine the nucleoid boundaries. As expected ParB-GFP foci colocalized exclusively with the Hoechst stain, and were equally spaced over the nucleoid (Fig. 1B,C,D for  $n_p = 1, 2$ , S2A,B Fig. for  $n_p = 3, 4$ ).

### Mathematical analysis shows that dynamic, competitive ParA concentrations can generate equal plasmid spacing

Several studies have proposed that plasmid positioning is controlled by a concentration gradient of ParA over the nucleoid [8,16–20]. Intuitively in this mechanism, ParB bound to plasmid *parC* (ParB-*parC* complex) interacts with nucleoid associated ParA-ATP, which effectively anchors the plasmid to the nucleoid. At the same time, the ParB-*parC* complex stimulates ParA-ATP hydrolysis causing a local ParA-ATP depletion. These processes could then generate a ParA-ATP gradient which a plasmid is able to follow. Reorganization of ParA gradients under the influence of multiple ParB-*parC* complexes might then lead to equal plasmid spacing. To rigorously understand if, and with what requirements, equal spacing can be achieved we develop here a minimal mathematical model based on the above principles.

We model the nucleoid as a 1d system of length  $L$  (along the long axis of the cell) on which ParA-ATP and plasmids can interact. Let  $A(x, t)$  denote the nucleoid-associated ParA-ATP concentration at position  $x$  relative to one nucleoid edge at time  $t$ . Let  $x_1(t) \dots x_{n_p}(t)$  be the positions of the  $n_p$  plasmids. ParA can bind to the nucleoid with flux  $J$ . Once bound, ParA-ATP can diffuse along the nucleoid with diffusion constant  $D$ . For simplicity, we first assume that the ParA-ATP concentration at each plasmid is zero due to a high ParA-ATP hydrolysis rate. Later on we will relax this assumption. This system can be described by the deterministic reaction-diffusion equations:

$$\frac{\partial A(x, t)}{\partial t} = D \frac{\partial^2 A(x, t)}{\partial x^2} + \frac{J}{L}$$

Boundary Conditions :  $A(x_i(t)) = 0$  for  $1 \leq i \leq n_p$

$$\left. \frac{\partial A(x, t)}{\partial x} \right|_{x=0} = 0 = \left. \frac{\partial A(x, t)}{\partial x} \right|_{x=L} \text{ for all } t.$$

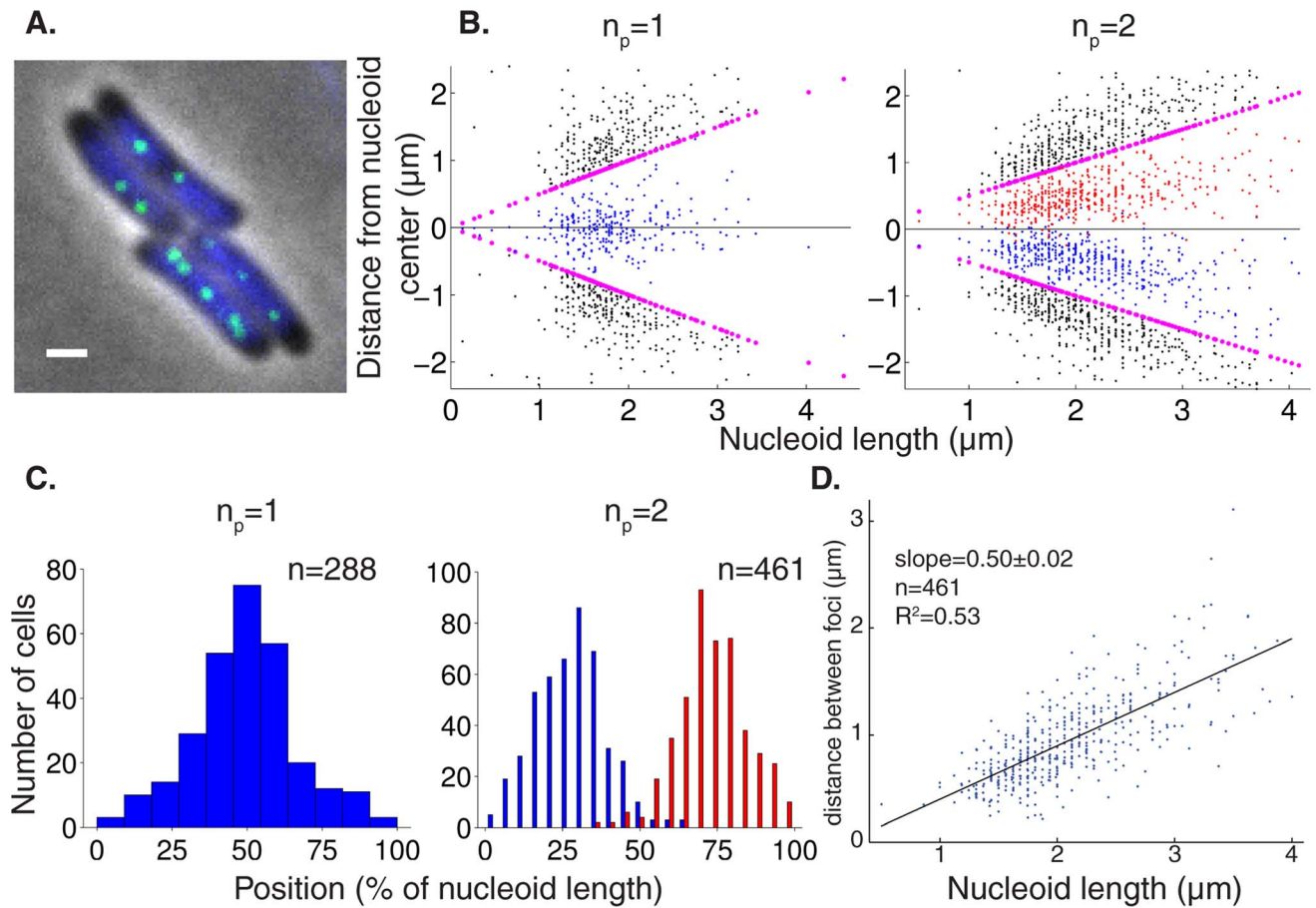
We now use separation of time scales to obtain the steady-state solution for  $A(x)$ : we assume that plasmid motion is much slower than the time for individual ParA-ATP molecules to diffuse over the nucleoid and generate a concentration profile. In this way, the plasmid positions  $x_1 \dots x_{n_p}$  are effectively time-independent and *a priori* unknown. The equation for  $A(x)$  then simplifies to:

$$\frac{d^2 A(x)}{dx^2} = - \frac{J}{LD}$$

Boundary Conditions :  $A(x_i) = 0$  for  $1 \leq i \leq n_p$  (1)

$$\left. \frac{dA(x)}{dx} \right|_{x=0} = 0 = \left. \frac{dA(x)}{dx} \right|_{x=L}$$

This equation can be solved by integrating twice using the boundary conditions. The solution is given by:



**Fig. 1. Plasmid foci are equally spaced over the nucleoid.** (A) Fluorescence localization of plasmid-binding protein ParB-GFP (green) and Hoechst DNA stain (blue) in representative WT *E. coli* cells. Scale bars: 1  $\mu\text{m}$ ; plasmid: pFS21 (mini-R1, *parC1*<sup>+</sup>, *parA*<sup>+</sup>, *parB::sfGFP*, *parC2*<sup>+</sup>). (B) Scatter plot of plasmid foci positions (blue, red) with respect to nucleoid edges (purple) and cell edges (black) for wild-type cells with  $n_p = 1, 2$  plasmid foci. (C) Histograms of plasmid foci positions shown in (B) relative to nucleoid length. (D) Scatter plot (blue) of the interplasmid focus distance as a function of nucleoid length in cells exhibiting two plasmid foci. A least square fit (black line) indicates a slope of 0.5.

doi:10.1371/journal.pcbi.1004009.g001

$$A(x) = \begin{cases} \frac{J}{2LD} (x_1^2 - x^2) & 0 \leq x \leq x_1 \\ \frac{J}{2LD} (-x^2 + (x_i + x_{i+1})x - x_i x_{i+1}) & x_i \leq x \leq x_{i+1} \text{ and } 1 \leq i < n_p \\ \frac{J}{2LD} [(L - x_{n_p})^2 - (L - x)^2] & x_{n_p} \leq x \leq L \end{cases} \quad (2)$$

Next we use these equations to compute the diffusive fluxes of ParA-ATP,  $j_i = |D \frac{dA}{dx}|$ , at a plasmid location  $x_i$ , where the + and - superscripts below refer to the flux from the right (+) and left (-) respectively. We find:

$$\begin{aligned} j_1^- &= \frac{Jx_1}{L}, \\ j_i^+ &= j_{i+1}^- = \frac{J(x_{i+1} - x_i)}{2L}, \\ j_{n_p}^+ &= \frac{J(L - x_{n_p})}{L} \end{aligned}$$

Clearly, a symmetric ParA concentration profile, where fluxes from either side balance, is only possible for  $x_1 = L - x_{n_p} = \frac{1}{2}(x_{i+1} - x_i)$ . The plasmids are then equally distributed with  $x_j = \frac{L}{2n_p} + \frac{L}{n_p}(j-1)$ . We note that the predicted interplasmid spacing  $\frac{L}{n_p}$  arising from this analysis is consistent with our experimental findings (Fig. 1D, S2B).

Importantly, the above analysis provides insight into the equal spacing mechanism. The key is that the above fluxes depend on the distances either between the plasmid and nucleoid end, or between neighboring plasmids. This feature is a consequence of ParA binding to the nucleoid anywhere, but with ParA release only occurring at a plasmid. In order for these on and off fluxes to balance at steady-state, the off-flux at a plasmid must scale with the inter-plasmid or plasmid-nucleoid-end distance. In this way, non-local information about lengths is converted into local spacing information encoded in the slope of ParA-ATP concentration. For non-equal plasmid spacing, the competing ParA concentrations on either side of a plasmid will be unequal, with one gradient steeper than the other. The steeper gradient corresponds to the side with

the greater available space for ParA binding. If a plasmid can preferentially move (on the appropriate slow time scale) towards the side with the locally steepest ParA-ATP concentration, the plasmids are then progressively restored towards equal spacing. As this process occurs, the ParA-ATP concentrations will dynamically reorganize such that a symmetric configuration around a plasmid is reached only when the plasmids are equally spaced. In this state, where the competing ParA-ATP concentrations are symmetric, plasmid movement would no longer have a directional preference and would thus remain, on average, stationary.

So far, we have assumed that the ParA-ATP concentration vanishes at a plasmid, corresponding to very fast ParA-ATP hydrolysis. However, our results also hold true when we only assume that this hydrolysis occurs with a finite rate  $k_B$ , leading to a non-zero concentration of ParA-ATP at a plasmid. This ParA-ATP can then anchor a plasmid to the nucleoid before being hydrolysed. This more general and realistic case is presented in the S1 Text, but our overall conclusions reached above remain unchanged.

From the above analysis, we see that the following conditions are required for equal plasmid spacing: (1) movement of a plasmid towards higher ParA-ATP concentrations. (2) diffusion of (at least a fraction of) ParA-ATP over the nucleoid to ensure formation of competitive concentration gradients. Single molecule tracking experiments *in vitro* support this assumption [17,18]. (3) ParA-ATP hydrolysis must occur (predominantly) by plasmid-associated ParB-*parC* complexes, again to ensure gradient formation. (4) ParA-ATP must adopt a 1d-like configuration, as previously claimed [9,13,14]. If ParA were not organized in this fashion, it would be possible for ParA to diffuse around the sides of a plasmid without encountering the hydrolyzing effect of the ParB-*parC* complex. This would equalize the ParA concentrations on both sides even in the case of asymmetrically placed plasmids, leading to failure of the equal spacing mechanism. This assumption is in line with our subsequent experiments (see below). Due to this proposed 1d-like nature, we will from now on refer to the ParA distributions away from a plasmid as ParA structures. (5) There must be a separation of time scales between plasmid movement and ParA concentration reorganization, as discussed above.

Importantly, this overall mechanism is not reliant on a specific type of plasmid translocation. Any process that would allow a plasmid to move into regions of higher ParA concentration will suffice. In the following sections we therefore analyze different means of plasmid movement and compare them with our experimental data to determine which is used in our *par2* segregation system.

### Diffusion/immobilization model could space highly mobile plasmids equally over the nucleoid

In the previous section the mechanistic details of plasmid movement towards a higher ParA concentration were not specified. We now examine a specific implementation involving a diffusion-immobilization mechanism. Using a minimal modeling approach, we assume that nucleoid-associated ParA-ATP can immobilize freely diffusing plasmids through its interaction with the ParB-*parC* complex and that ParA-ATP does not polymerize (Fig. 2A). Since the plasmid will tend to become immobilized in regions of higher ParA-ATP concentration, this process allows for effective plasmid translocation up a ParA-ATP concentration gradient. We also incorporate ParB-*parC*-stimulated ParA-ATP hydrolysis at a plasmid, in accordance with prior experimental data. To further investigate this mechanism, given the known physiological and biochemical constraints, we developed stochastic simulations using a Gillespie algorithm [22]. Here we use standard

diffusion for the plasmid movement; below we discuss the potential impact of subdiffusive motion.

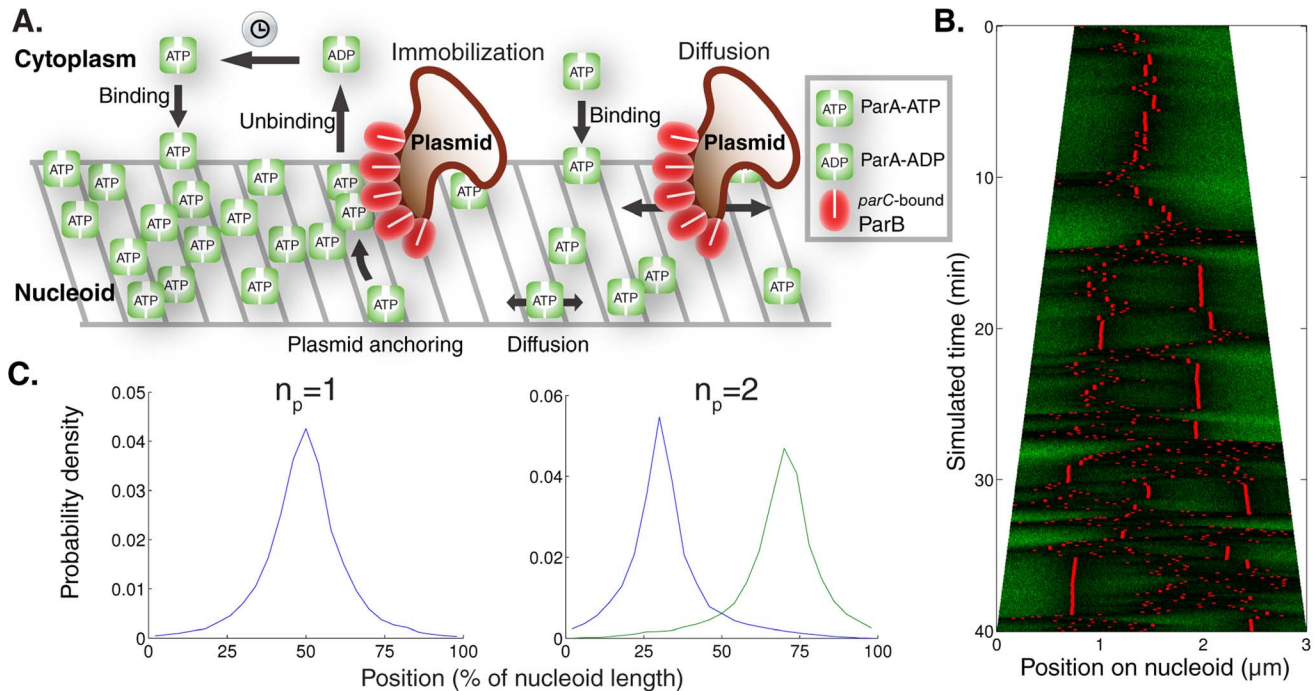
In our simulation, a one dimensional lattice with sites of size  $dx = 5$  nm represents the nucleoid. ParA-ATP and plasmids can diffuse on the lattice with diffusion coefficient  $D_A$  and  $D_P$  respectively. Up to 35 ParA-ATP can bind to a plasmid at the same site with reaction parameter  $k_{AB}$  reflecting the binding interaction of ParA-ATP and the ParB-*parC* complex [11]. More than one ParA-ATP bound to a plasmid reduces the plasmid diffusion constant to zero. Plasmid-bound ParA-ATP can be hydrolysed with reaction parameter  $k_B$ . Whenever a ParA-ATP hydrolysis event occurs, ParA unbinds from the nucleoid and becomes a cytoplasmic ParA-ADP. ParA-ADP can then be converted into a cytoplasmic ParA-ATP that is competent in DNA binding (cytoplasmic ParA-ATP for short) with a slow reaction parameter  $k_W$  [8]. Cytoplasmic ParA-ATP can then bind anywhere along the nucleoid with parameter  $k_{on}$  (see Materials and Methods and Tables 1,2 for details).

Prior work has demonstrated plasmid displacement along the long cell axis of up to 3–4  $\mu\text{m}$  within 10 min [9,15]. With a diffusion/immobilization mechanism all plasmid movement in between immobilization events is generated by (unbiased) free diffusion, for which we have (in 1d) a mean square displacement (MSD) of  $\langle r^2(t) \rangle = 2D_P t$ . By inserting the above length and time scales into this equation, we conclude that a plasmid diffusivity of at least  $D_P \sim 10^{-2} \mu\text{m}^2\text{s}^{-1}$  would be required to generate sufficiently rapid diffusive movement in accordance with previous experiments. We therefore chose  $D_P = 10^{-1} \mu\text{m}^2\text{s}^{-1}$ . In order to physically justify that ParA can immobilize the plasmids, we chose the nucleoid bound ParA-ATP diffusivity to be lower than  $D_P$ , with  $D_A = 10^{-2} \mu\text{m}^2\text{s}^{-1}$  (Table 2). We experimentally constrained the overall copy number of ParA for pB171 *par2* by semi-quantitative Western blots, which revealed that there were approximately  $8 \times 10^3$  ParA monomers per cell (S1B Fig.). This diffusion/immobilization model could produce equal plasmid spacing on simulated growing nucleoids with varying numbers of plasmids (Fig. 2B,C, S3A). This result demonstrates that using a sufficiently high (low) plasmid (ParA) diffusivity, respectively, the equal plasmid spacing seen in our experiments (Fig. 1B,C,D, S2A,B Fig.) and previously [9], could in principle be achieved using a diffusion/immobilization mechanism.

### Free plasmid mobility is too low for a diffusion/immobilization mechanism

To test whether the requirement of a relatively high free plasmid mobility is met *in vivo*, we compared the movement of test-plasmids with and without *par2*. We analyzed trajectories of labeled plasmid foci using the *tetO*-TetR-mCherry labeling system, measuring the positions over time (Fig. 3A) and MSDs for each time lag  $\tau$ . Plasmid motion will be biased by a functional *par2*<sup>+</sup> partitioning system, in contrast to the random motion of *par*<sup>−</sup>. Nevertheless comparing MSDs can still be informative in comparing relative overall mobilities. On time scales up to a minute we found that the *par2*<sup>+</sup> MSD is higher than in *par*<sup>−</sup> (Fig. 3B), showing that, on average, *par2*<sup>+</sup> plasmids are more mobile than their *par*<sup>−</sup> counterparts. Note that the number of data points for the short time lags far exceeds the number of trajectories ( $n_{\text{par}^-} = 747$ ,  $n_{\text{par2}^+} = 763$ ), since every trajectory contains multiple short time lags. Consequently our estimates for the mean are relatively precise for short time lags. It is true that the error on the mean does not reflect inaccuracy due to experimental limitations in determining the actual plasmid position, for instance due to a finite pixel size. However, that error is the same for both *par2*<sup>+</sup> and *par*<sup>−</sup>. Moreover, since the error is also time lag independent, it





**Fig. 2. Diffusion/immobilization model can move and maintain plasmids at equally spaced positions. (A)** Schematic illustration of *par2* diffusion/immobilization model. The clock indicates the slow conversion of cytoplasmic ParA-ADP into cytoplasmic ParA-ATP that is competent to bind to the nucleoid. **(B)** Typical simulation kymograph of diffusion/immobilization model for growing cell, where plasmid (red) diffusion influenced by the local ParA-ATP (green) concentration leads to immobilization initially at mid cell. After plasmid duplication, the system dynamically self-organizes to reacquire equal plasmid spacing. **(C)** Time-averaged plasmid position distributions for diffusion/immobilization model with  $n_p = 1-2$  on a simulated nucleoid growing from 1.5 μm to 3 μm in 40 min without plasmid duplication. Plasmid distributions were obtained by sampling positions every 5 s in 36 independent simulations. doi:10.1371/journal.pcbi.1004009.g002

is taken into account in our fitting procedure as a time lag independent term (for more details see below and Materials and Methods). Overall, these results are hard to reconcile with a diffusion/immobilization mechanism where the *par2* system can

only immobilize plasmids, and thus lower their MSD. These MSD values could in principle be limited due to cellular confinement. However, we found that MSD saturation only starts to occur at much larger length scales at times of up to 10 min (Fig. 3C). In the

**Table 1. Reactions and propensities used in the diffusion/immobilization model.**

Reactions	Propensities $p_i$
$A_i \rightarrow A_{i+1}, i = 0 \dots (L-2)$	$\frac{D_A}{dx^2} \cdot A[i]$
$A_{i+1} \rightarrow A_i, i = 0 \dots (L-2)$	$\frac{D_A}{dx^2} \cdot A[i+1]$
$P_{j,i} \rightarrow P_{j,i+1}, i = 0 \dots (L-2), j = 0 \dots 35$	$\frac{D_P}{dx^2} \cdot P[j][i], \text{ if } j = 0$ $\frac{D_A}{dx^2} \cdot P[j][i], \text{ if } j = 1$ 0, if $j > 1$
$P_{j,i+1} \rightarrow P_{j,i}, i = 0 \dots (L-2), j = 0 \dots 35$	$\frac{D_P}{dx^2} \cdot P[j][i], \text{ if } j = 0$ $\frac{D_A}{dx^2} \cdot P[j][i], \text{ if } j = 1$ 0, if $j > 1$
$P_{j,i} + A_i \rightarrow P_{j+1,i}, i = 0 \dots (L-1), j = 0 \dots 34$	$k_{AB} \cdot P[j][i] \cdot A[i]$
$P_{j+1,i} \rightarrow P_{j,i} + A_{ADP}, i = 0 \dots (L-1), j = 0 \dots 34$	$k_B \cdot P[j+1][i]$
$A_{ADP} \rightarrow A_{cyto}$	$k_W \cdot A_{ADP}$
$A_{cyto} \rightarrow A_i, i = 0 \dots (L-1)$	$k_{on}/L \cdot A_{CYTO}$

doi:10.1371/journal.pcbi.1004009.t001

**Table 2.** Parameter values used in the diffusion/immobilization model.

Parameter	Description	Value	Notes
$D_A$	Nucleoid bound ParA-ATP diffusion constant	$10^{-2} \mu\text{m}^2/\text{s}$	Fitted, can be increased without loss of qualitative behaviour of system. Nevertheless, it is difficult to physically reconcile more mobile nucleoid-bound ParA-ATP with the ability to immobilize a plasmid with a lower diffusion constant. Therefore we have assumed that ParA-ATP diffuses 10x slower than the plasmid, ensuring that the assumption that ParA-ATP can immobilize plasmids is physically justified.
$D_P$	Plasmid diffusion constant	$10^{-1} \mu\text{m}^2/\text{s}$	A relatively high value is needed for compatibility with previous experiments [9]. An upper bound on the plasmid diffusion constant from experiments (Fig. 3C) turned out to be too low for this model to fit our experimental observations.
$k_{on}$	ParA-ATP nucleoid binding	$50 \text{ s}^{-1}$	Constrained by experiment [8].
$k_{AB}$	ParA-ATP to plasmid binding	$100 \text{ s}^{-1}$	Fitted, should be high enough to allow for plasmid immobilization.
$k_B$	Plasmid bound ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$68.5 \text{ s}^{-1}$	Fitted together with $D_A$ and $k_W$ to ensure equal plasmid spacing.
$k_W$	(Cytoplasmic) ParA-ADP to ParA-ATP conversion	$1/15 \text{ s}^{-1}$	Constrained by experiment [8], this value should be low enough to ensure that cytoplasmic ParA diffusion can generate a uniform cytoplasmic ParA-ATP and ParA-ADP concentration.

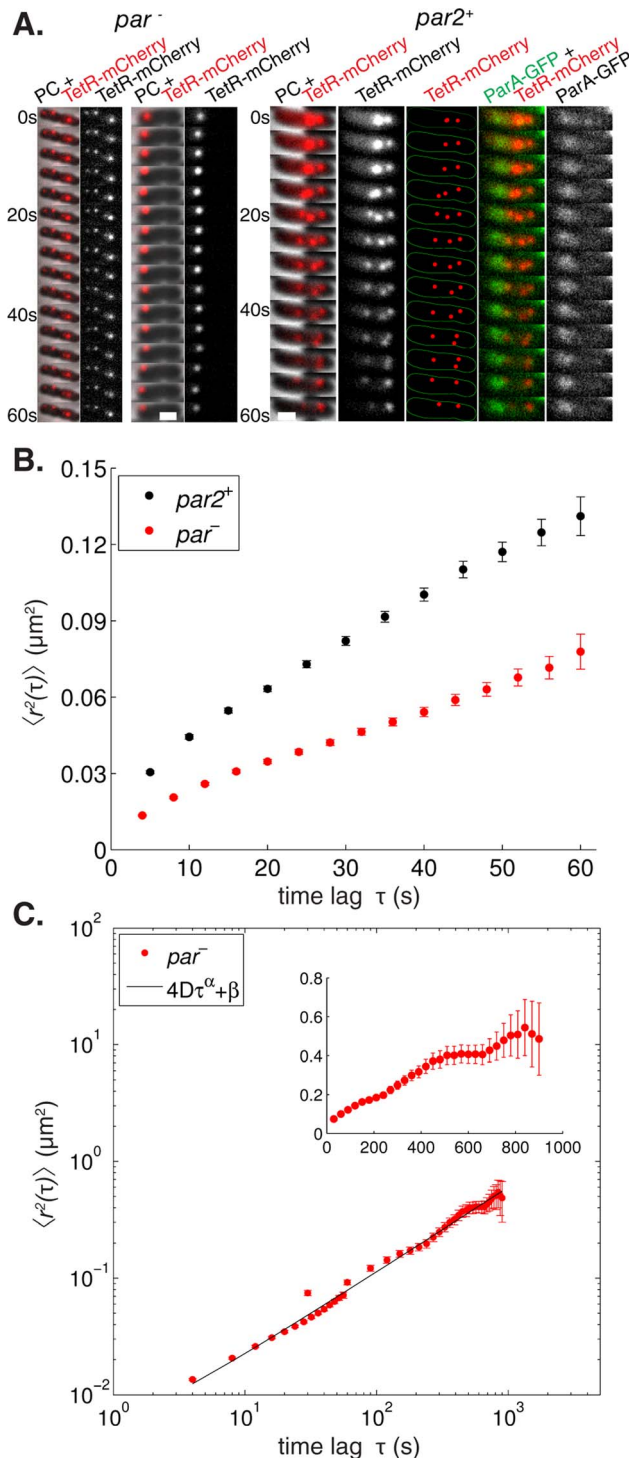
doi:10.1371/journal.pcbi.1004009.t002

presence of *par2*, plasmids generally reside within the nucleoid region, while in its absence they tend to become somewhat more polar localized, although they can still sample the entire cell volume on long enough timescales [23]. Consistently we still find many *par<sup>-</sup>* plasmids located within the nucleoid region (S3B Fig.). Restricting the mobility analysis to *par<sup>-</sup>* plasmids within the nucleoid region did not alter the resulting MSD curves significantly (S3B Fig.). We conclude that the presence of *par2* can increase plasmid mobility in the nucleoid region, which is inconsistent with a diffusion/immobilization mechanism. We emphasize that this conclusion can be made irrespective of the underlying (*par<sup>-</sup>*) plasmid transport processes, which we now describe in more detail.

It has been reported that chromosomal loci and RNA-protein particles exhibit subdiffusive, rather than diffusive, behavior in the cytoplasm [24,25]. Therefore it is possible that plasmids without a segregation mechanism could also exhibit subdiffusive motion. Further analysis is required to fully distinguish subdiffusion from the additional effects of cellular confinement or glass-like properties of the bacterial cytoplasm [23,25]. Nevertheless such additional analysis is not required for the conclusions on *par<sup>-</sup>* plasmid mobility relevant to this study, as we now explain. Subdiffusion results in an expected MSD displacement of the form  $\langle r^2(\tau) \rangle = 4D\tau^\alpha$ , with  $\alpha < 1$  and  $D$  the apparent diffusion constant (in units of  $\mu\text{m}^2\text{s}^{-\alpha}$ ). We find that our MSD displacements on both short and long timescales are well described by subdiffusion with  $\alpha = 0.7\text{--}0.8$  and an apparent diffusion constant  $D = 5\text{--}10 \times 10^{-4} \mu\text{m}^2\text{s}^{-\alpha}$  (Fig. 3C and Materials and Methods for details). This is consistent with other recent reports on *par<sup>-</sup>* plasmid mobility [23,26]. Importantly the experimental MSD is lower on all observed timescales than a hypothetical particle that would perform free diffusion inside a cell with a diffusion constant  $D_f = 10 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ . This upper limit is already much lower than that needed to be consistent with the previously reported plasmid displacement data discussed above. We will further exploit this upper limit in our analysis below.

To further investigate the effect of *par2* on plasmid positioning, we also studied rapid plasmid segregation events. We defined these as cases where two plasmid foci whose separation is initially  $\leq 0.3 \mu\text{m}$ , move within 20 s at least another  $0.8 \mu\text{m}$  apart (Fig. 3A,

S3C). We also allowed for the two foci to be initially merged. Using these criteria, despite equally large data sets, we found 13 such events in *par2<sup>+</sup>* and only one such case in *par<sup>-</sup>*. Furthermore, we only retrieved 2 further *par2<sup>+</sup>* segregation events when we relaxed the criterion to separation within 60 s instead of 20 s. This analysis shows that most segregation events occur rapidly. When we investigated the 26 plasmid trajectories involved they showed larger maximal MSDs compared to sets of 26 trajectories that were repeatedly randomly sampled from the whole *par2<sup>+</sup>* dataset ( $p < 10^{-6}$ ). This finding indicates that the *par2* system can particularly enhance the mobility of plasmids when they are in close proximity. We then simulated 300 plasmid duplication events with our diffusion/immobilization model to determine the magnitude of diffusion constant required to generate the experimentally observed segregation. Note that we used diffusion rather than subdiffusion here because we have already determined that *par<sup>-</sup>* plasmid movement is slower on all observed timescales than free diffusion with a diffusion constant  $D_f = 10 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ . Hence, if the required diffusion constant is larger than  $D_f$  then we have also ruled out a subdiffusion/immobilization model. We required that 5% (15 out of 300) of segregated distances within 20 s were at least  $0.8 \mu\text{m}$  (a very conservative requirement, since the criterion was satisfied by 13 of our 15 experimental segregation events). This requirement necessitated a free plasmid diffusion constant on the order of  $10^{-1} \mu\text{m}^2\text{s}^{-1}$ , about two orders of magnitude higher than our experimentally observed upper bound  $D_f$  on the experimental *par<sup>-</sup>* plasmid mobility. Hence, we conclude that the plasmids are generally too immobile for a diffusion/immobilization (or subdiffusion/immobilization) mechanism to explain these segregation events. Also the qualitative behaviour of segregation events in the diffusion/immobilization model appears different, since experimental segregation events (Fig. 3A, S3C) show more directionally biased motion, while the diffusion/immobilization model generates more sustained random, diffusive motion during segregation, prior to immobilization at equally spaced positions (Fig. 2B). Nevertheless, these segregation events were sufficiently rare not to significantly alter the overall MSD behaviour of the entire dataset shown in Fig. 3B. Thus the increased average mobility in the presence of *par2<sup>+</sup>* cannot only be ascribed to these segregation events.



**Fig. 3. The *par2* segregation system increases plasmid mobility.** (A) Time lapses showing the localization of *par*<sup>-</sup> pMH82tetO120 (mini-R1, *par*<sup>-</sup>, *tetO120*) and *par2*<sup>+</sup> pSR236 (mini-R1, *parC1*<sup>+</sup>, *parA*<sup>+</sup>, *parB*<sup>+</sup>, *parC2*<sup>+</sup>, *tetO120*, *P<sub>lac</sub>::parA::eGFP*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). The *par2*<sup>+</sup> time lapse, with ParA-GFP localization, shows a segregation event where two foci segregate  $\geq 0.8 \mu\text{m}$  further apart within 20 s. PC = phase contrast, scale bar: 1  $\mu\text{m}$ . (B) Mean square displacements  $\langle r^2(\tau) \rangle$  after time lag  $\tau$  were extracted from plasmid trajectories ( $n_{\text{par}^-} = 747$ ,  $n_{\text{par2}^+} = 763$ ) using strains specified in (A), *par*<sup>-</sup> (red) and *par2*<sup>+</sup> (black), error bars: standard error of the mean. (C) Log-log plot of experimental mean square displacements  $\langle r^2(\tau) \rangle$  after time lag  $\tau$  (red) were extracted from plasmid trajectories over

1 min as in (A,B) and (inset, linear scales,  $n = 50$ ) over 15 min from *par*<sup>-</sup> pMH82tetO120 (mini-R1, *par*<sup>-</sup>, *tetO120*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). At timescales on the order of 10 min saturation of the MSD occurs due to cellular confinement. A nonlinear least square fit (black line) using the function  $\langle r^2(\tau) \rangle = 4D\tau^\alpha + \beta$  was used to estimate parameter values:  $\alpha = 0.73 \pm 0.02$ ,  $D = 9.7 \pm 1.3 \times 10^{-4} \mu\text{m}^2\text{s}^{-\alpha}$ ,  $\beta = 1.6 \pm 2.4 \times 10^{-3} \mu\text{m}^2$ , ( $R^2 = 0.99$ , p-values:  $8 \times 10^{-15}$ ,  $8 \times 10^{-3}$  and 0.50 respectively). See Materials and Methods for details; error bars: standard error of the mean.  
doi:10.1371/journal.pcbi.1004009.g003

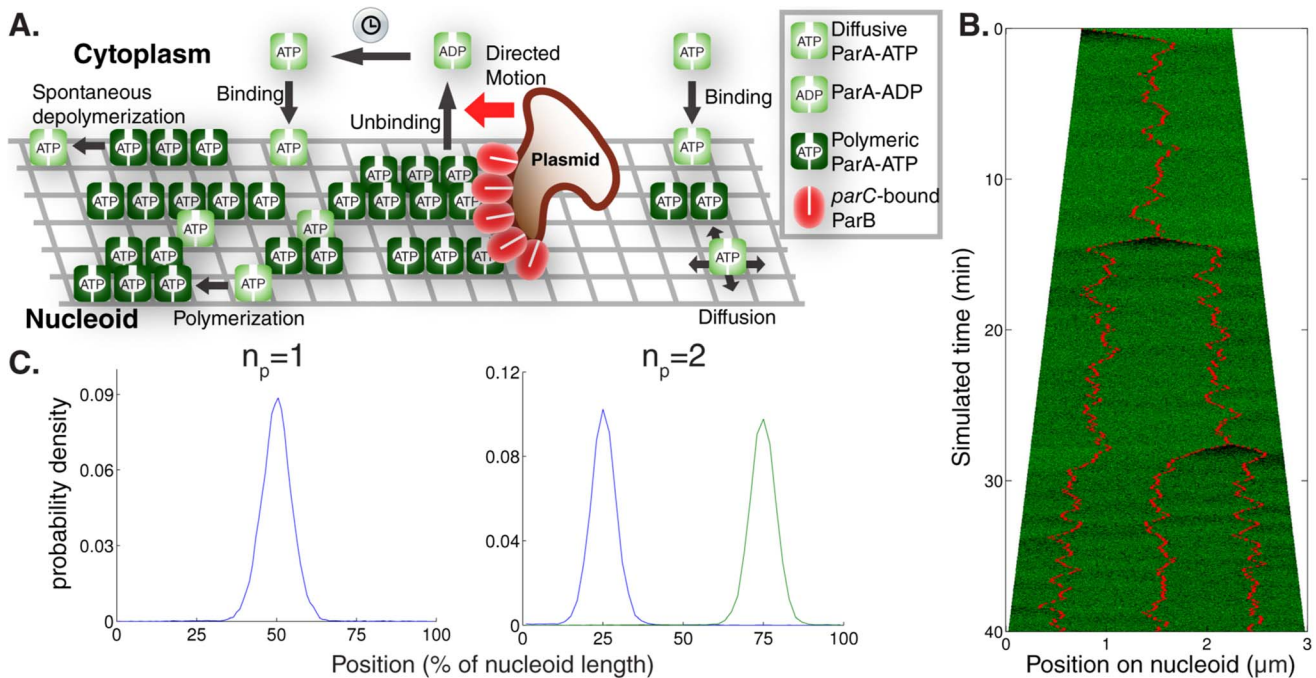
It is possible that the *tetO*-TetR-mCherry labeling system caused reduced plasmid mobility as compared to unlabelled plasmids. However, as we used the same labeling method for both *par2*<sup>+</sup> and *par*<sup>-</sup> cases, our above conclusions on relative mobility are unaffected. Moreover, our *tetO*-TetR-mCherry labeled plasmids still exhibited rapid segregation events (such as in Fig. 3A), underscoring the ability of *par2* to overcome low plasmid mobility. Overall, we find that diffusion/immobilization cannot explain our data on *par2*<sup>+</sup> versus *par*<sup>-</sup> plasmid mobility, as well as on rapid *par2*<sup>+</sup> plasmid segregation.

### ParA structures competing to direct plasmid motion can space plasmids equally over the nucleoid

Given the shortcomings of the diffusion/immobilization model, we next tested models based on directed motion, allowing more rapid directed rather than unbiased diffusive plasmid movement. More specifically, we tested models based on the formation of competing ParA polymers, with ParB-*parC*-stimulated ParA-ATP hydrolysis directing plasmid movement. By modulating the length of these polymers, we thereby tested the robustness of directed motion models to generate equal plasmid positioning.

We again used a Gillespie algorithm to simulate ParA dynamics on the nucleoid (see Fig. 4A, Materials and Methods and Tables 3,4 for details). The nucleoid was represented as a rectangular lattice ( $dx = 5 \text{ nm}$  in both dimensions), with a much shorter width (30 nm) than length (several  $\mu\text{m}$ ). Similar reactions as in the diffusion/immobilization model described the cytoplasmic dynamics of ParA-ADP and ParA-ATP. Nucleoid-associated ParA-ATP could also still diffuse across the nucleoid in a mobile state in all four directions to neighbouring sites with diffusion constant  $D_A$ . However, two of these molecules at sites neighboring each other along the long nucleoid axis could interact to form a ParA polymer of two subunits, with reaction parameter  $k_p$ . Further ParA-ATP polymerization could occur by attachment of mobile ParA-ATP, located at a site immediately next to the tip of an existing ParA polymer, but only along the long axis. ParA-ATP polymers were assumed to be immobile. A ParA-ATP polymeric subunit could depolymerize spontaneously with reaction parameter  $k_{dp}$ , i.e. be converted into a mobile ParA-ATP at the same site. Given that its size is similar to the width of the lattice, we only took into account the plasmid position along the long axis and we assumed that it occupied all sites along the short axis simultaneously. The plasmid could diffuse with our experimentally estimated diffusion coefficient  $D_P$  along the long axis when polymeric ParA-ATP was not present either at any of the sites that the plasmid occupied or sites neighbouring the plasmid. In the presence of polymeric ParA-ATP, the plasmid was assumed to be tethered to such a polymer (via a ParB-*parC* complex), which prevented plasmid diffusion. At sites with a plasmid present, polymeric ParA-ATP could be converted into cytoplasmic ParA-ADP with reaction parameter  $k_B$ . Reflecting directed motion, at sites neighbouring a plasmid occupied by polymeric ParA-ATP, a plasmid could with reaction parameter  $k_{dm}$  move to the coordinate





**Fig. 4. The directed motion model can move and maintain plasmids at equally spaced positions.** (A) Schematic illustration of *par2* directed motion model. The clock indicates the slow conversion of cytoplasmic ParA-ADP into cytoplasmic ParA-ATP that is competent to bind to the nucleoid. (B) Typical simulation kymograph of directed motion model with short polymers for a simulated growing cell where a plasmid (red) is initially directed from a nucleoid edge to mid-cell by ParA (green) filament competition. After plasmid duplication, the system dynamically self-organizes to attain equal spacing. (C) Time-averaged plasmid position distributions for directed motion model with short polymers for  $n_p = 1, 2$  plasmids on a simulated nucleoid growing from 1.5 μm to 3 μm in 40 min without plasmid duplication. Plasmid distributions were obtained by sampling positions every 5 s in 36 independent simulations.  
doi:10.1371/journal.pcbi.1004009.g004

along the long axis of that ParA-ATP subunit, coinciding with conversion of that ParA-ATP into cytoplasmic ParA-ADP. For wild-type simulations, any plasmid in the system formed a hard wall to mobile ParA-ATP diffusion so that diffusing ParA-ATP molecules could not diffuse past a plasmid.

We first adjusted the ParA-ATP polymerization rate to generate short filaments, of approximately 10 subunits in length (Table 4 for parameters). Simulations again faithfully reproduced the equal spacing of plasmids along simulated growing nucleoids with varying numbers  $n_p$  of plasmids (Fig. 4B,  $n_p = 1, 2$  in Fig. 4C,  $n_p = 3, 4$  in S4A Fig.) in good agreement with our experiments (Fig. 1, S2). By adjusting the ParA-ATP polymerization rate (Table 4), long continuous ParA polymer bundles could also be generated. In that case equal spacing could also be achieved (S4B,C Fig.). Intuitively, in both short and long filament cases, this occurs because in an irregularly spaced plasmid configuration, the unequal ParA concentrations on either side of a plasmid result in an unequal degree of ParA polymerization. This in turn results in an unequal amount of competitive directed motion events to each side, resulting in effective directed translocation over longer length scales back towards an equally positioned state. Plasmid separation occurs when two nearby plasmids encounter two ParA-ATP structures extending in opposite directions away from the plasmids. The two ParA-ATP structures will then necessarily mediate a segregation event. The effect of directed movement in this model is clearest in the case of plasmid segregation events (Fig. 4B, S4B), where we see rapid segregation consistent with the fast segregation events observed experimentally (see Fig. 3A).

### ParA-GFP oscillations are not continuously required for equal plasmid spacing

Intriguingly, simulations of the directed motion model did not generally produce sustained spatiotemporal oscillations of ParA across the nucleoid (short polymers: Fig. 4B, long polymers: S4B Fig.). A lack of sustained oscillations would therefore appear to be a common feature of models where competitive ParA structures generate equal plasmid spacing. This absence was unexpected, as prior experimental work had emphasized the oscillatory aspect of the ParA dynamics [12–14]. To experimentally test this key model prediction in an unbiased fashion, we experimentally measured the degree of ParA asymmetry in the *par2* system in a large dataset ( $n = 134$ ) of snapshots of ParA-GFP across the nucleoid. We examined only cases with a single plasmid *tetO*-TetR-mCherry focus, where sustained oscillations should be easiest to infer. The ParA-GFP fluorescence signal from pole to plasmid position was summed and divided by the respective pole-to-plasmid distance. This generated two ParA-GFP fluorescence densities  $I_L$  and  $I_R$  for either side extending to the two cell poles. This allowed us to compute the normalized asymmetry measure  $|I_L - I_R| / |I_L + I_R|$  [27] for ParA (see Materials and Methods for details). Asymmetric ParA-GFP distributions, arising for example from oscillations, where for example  $I_L \approx 0$ ,  $I_R \approx 1$ , will give asymmetry values closer to one, whereas symmetric ParA-GFP distributions, where  $I_L \approx I_R$ , will give values closer to zero. Note that the ParA-GFP exposure time used here was 1.5 s; clearly, we cannot measure asymmetries that occur on a timescale faster than this exposure time. However, the timescales of the plasmid and ParA-GFP dynamics are on the order of tens of seconds or longer and it is therefore unlikely

**Table 3.** Reactions and propensities used in the directed motion models.

Reactions	Propensities $p_t$
$Am_{ij} \rightarrow Am_{i+1,j}, i=0 \dots (L-2), j=0 \dots (S-1)$	$\frac{D_A}{dx^2} \cdot A_m[i][j], \text{ if } P[i+1] = 0$ 0, otherwise
$Am_{i+1,j} \rightarrow Am_{i,j}, i=0 \dots (L-2), j=0 \dots (S-1)$	$\frac{D_A}{dx^2} \cdot A_m[i+1][j], \text{ if } P[i] = 0$ 0, otherwise
$Am_{ij} \rightarrow Am_{i,j+1}, i=0 \dots (L-1), j=0 \dots (S-2)$	$\frac{D_A}{dx^2} \cdot A_m[i][j]$
$Am_{i,j+1} \rightarrow Am_{i,j}, i=0 \dots (L-1), j=0 \dots (S-2)$	$\frac{D_A}{dx^2} \cdot A_m[i][j+1]$
$P_i \rightarrow P_{i+1}, i=0 \dots (L-2)$	$\frac{D_P}{dx^2} \cdot P[i], \text{ if } \sum_{k=i,i \pm 1, j=0 \dots S-1} A[k][j] = 0$ 0, otherwise
$P_{i+1} \rightarrow P_i, i=0 \dots (L-2)$	$\frac{D_P}{dx^2} \cdot P[i], \text{ if } \sum_{k=i,i \pm 1, j=0 \dots S-1} A[k][j] = 0$ 0, otherwise
$A_{i,j} \rightarrow Am_{i,j}, i=0 \dots (L-1), j=0 \dots (S-1)$	$k_{dp} \cdot A[i][j], \text{ if } P[i] = 0$ 0, otherwise
$Am_{ij} + Am_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_p \cdot A_m[i][j] \cdot A_m[i+1][j], \text{ if } P[i] = P[i+1] = A[i][j] = A[i+1][j] = 0$ 0, otherwise
$A_{i,j} + Am_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_p \cdot A[i][j] \cdot A_m[i+1][j], \text{ if } P[i] = P[i+1] = A[i+1][j] = 0$ 0, otherwise
$Am_{ij} + A_{i+1,j} \rightarrow A_{i,j} + A_{i+1,j}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_p \cdot A_m[i][j] \cdot A[i+1][j], \text{ if } P[i] = P[i+1] = A[i][j] = 0$ 0, otherwise
$P_i + Am_{ij} \rightarrow P_i + A_{ADP}, i=0 \dots (L-1), j=0 \dots (S-1)$	$k_{mB} \cdot P[i] \cdot A_m[i][j]$
$P_i + Am_{i+1,j} \rightarrow P_i + A_{ADP}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_{mB} \cdot P[i] \cdot A_m[i+1][j]$
$P_{i+1} + Am_{ij} \rightarrow P_{i+1} + A_{ADP}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_{mB} \cdot P[i+1] \cdot A_m[i][j]$
$P_i + A_{i,j} \rightarrow P_i + A_{ADP}, i=0 \dots (L-1), j=0 \dots (S-1)$	$k_B \cdot P[i] \cdot A[i][j]$
$P_i + A_{i+1,j} \rightarrow P_{i+1} + A_{ADP}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_{dm} \cdot P[i] \cdot A[i+1][j]$
$P_{i+1} + A_{i,j} \rightarrow P_i + A_{ADP}, i=0 \dots (L-2), j=0 \dots (S-1)$	$k_{dm} \cdot P[i+1] \cdot A[i][j]$
$A_{ADP} \rightarrow A_{cyto}$	$k_W \cdot A_{ADP}$
$A_{cyto} \rightarrow Am_{i,j}, i=0 \dots (L-1), j=0 \dots (S-1)$	$k_{on} / (L \cdot S) \cdot A_{CYTO}$

doi:10.1371/journal.pcbi.1004009.t003

that any significant asymmetry is being missed by our measurements.

When we examined our whole distribution of cells exhibiting single plasmid *tetO*-TetR-mCherry foci, we found that the degree of ParA-GFP asymmetry (Fig. 5A,B) was low in comparison with the well-established MinD spatiotemporal oscillator [27]. Furthermore, the ParA-GFP asymmetry did not correlate with cell length (S5A Fig.,  $R^2 = 0.08$ ), unlike the case of MinD-YFP [27]. We also compared the ParA-GFP asymmetry to the Hoechst signal. This DNA stain labels the nucleoid itself, which is relatively uniform along the long cell axis [28–30]. Here, any asymmetry is not expected to depend on the plasmid foci positions. The Hoechst asymmetry distribution was indeed concentrated around relatively small values, but was apparently measurable within our approach (Fig. 5B, S5B). Importantly, we found that the ParA-GFP asymmetry measure had a similarly low value as for the Hoechst case (Fig. 5B, S5B, no significant difference, Kolmogorov-Smirnov test), and that for both the asymmetry is uncorrelated to the plasmid focus position (S5C Fig.). We therefore conclude that for a single plasmid focus, ParA-GFP typically resides on both sides of a plasmid, with relatively little asymmetry or oscillation, as predicted by the directed motion model, irrespective of a weak (Fig. 5B) or strong (S5B Fig.) degree of polymerization.

Previous analyses had focused on plasmids migrating in the wake of retracting ParA-GFP structures [9]. Such events can transiently give rise to relatively high ParA-GFP asymmetries (see, for example, Fig. 3A, 5C). Accordingly, we conclude that ParA asymmetry or oscillations are not continuously required for *par2* mediated plasmid positioning. Transient asymmetry, including oscillations, instead likely arises from the dynamics needed to bring about equal plasmid spacing following a spatial perturbation or plasmid duplication event (Fig. 5C). Once the ParA distribution has returned to being relatively symmetric, this coincides with an equally spaced plasmid configuration (Fig. 5C). Such dynamics can be seen in our model simulations (Figs. 4B, S4B): asymmetric during plasmid segregation events, but relatively symmetric otherwise. This analysis can therefore accommodate both our findings of a relatively symmetric ParA distribution with previous reports emphasizing asymmetry and oscillations. Overall, our finding of predominantly symmetric, non-oscillatory ParA dynamics may help to reconcile similar findings for ParA in other plasmid partitioning systems, such as for plasmid P1 [15,16].

### ParA-GFP forms structures within the nucleoid region

One required feature to achieve equal plasmid spacing is that the ParA-ATP should be organized in a 1d-like structure along the

**Table 4.** Parameter values used in the directed motion models.

Parameter	Description	Value	Notes
$D_A$	Nucleoid bound ParA-ATP diffusion constant	$1 \mu\text{m}^2/\text{s}$	Constrained by experiment [18], value can be varied by several orders of magnitude without loss of qualitative behaviour of system. Note that this form of ParA-ATP does not have an effect on the mobility of plasmids, since only polymeric ParA-ATP, immobile due to the interaction with the nucleoid, can direct the motion of a plasmid.
$D_P$	Plasmid diffusion constant	$3 \times 10^{-4} \mu\text{m}^2/\text{s}$	Constrained by experiment (Fig. 3C).
$k_{on}$	ParA-ATP nucleoid binding	$50 \text{ s}^{-1}$	Constrained by experiment [8].
$k_B$	Plasmid bound polymeric ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$68.5 \text{ s}^{-1}$	Chosen to be the same as $k_B$ in diffusion/immobilization model for consistency; constrained by $k_B \gg k_{dm}$ which ensures that all the ParA-ATP at the location of a plasmid is converted into cytoplasmic ParA-ADP before the plasmid moves to a neighboring site. Value can be varied within a wide range without loss of qualitative behaviour of system.
$k_{mB}$	Plasmid bound mobile ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$40 \text{ s}^{-1}$	Fitted, value can be varied within a wide range without loss of qualitative behaviour of system. Setting this rate too high depletes ParA-ATP locally around a plasmid, which inhibits directed plasmid motion events.
$k_W$	(Cytoplasmic) ParA-ADP to ParA-ATP conversion	$1/15 \text{ s}^{-1}$	Constrained by experiment [8], this value should be low enough to ensure that cytoplasmic ParA diffusion can generate a uniform cytoplasmic ParA-ATP and ParA-ADP concentration.
$k_{dm}$	Plasmid directed motion rate (in presence of one neighboring plasmid)	$0.8 \text{ s}^{-1}$	Constrained by experiment (Fig. 2A). If interpreted as biased plasmid diffusion along the polymer (burnt-bridge mechanism [4]), this would result effectively in a maximal plasmid diffusion constant of $1 \times 10^{-4} \mu\text{m}^2 \text{ s}^{-1}$ (short) and $1.2 \times 10^{-4} \mu\text{m}^2 \text{ s}^{-1}$ (long). These values are consistent with the free diffusion constant $D_P$ (see above), since the interaction with immobile ParA-ATP polymers could lower the plasmid mobility.
$k_p$	Polymerization: mobile ParA-ATP to polymeric ParA-ATP conversion	$800 \text{ s}^{-1}$ (short), $10^6 \text{ s}^{-1}$ (long)	Fitted together with $k_{dm}$ and $k_W$ to ensure equal plasmid spacing. $k_p$ and $k_{dp}$ together with the total ParA-ATP concentration determine the extent of ParA-ATP polymerization.
$k_{dp}$	Spontaneous depolymerization: ParA-ATP to mobile ParA-ATP conversion	$10 \text{ s}^{-1}$ (short), $10^{-4} \text{ s}^{-1}$ (long)	Fitted. See notes on $k_p$ parameter above.
$S$	Short axis length of the nucleoid region where nucleoid bound ParA-ATP can polymerize.	30 nm (short), 25 nm (long)	Fitted, values should be small compared to the long nucleoid axis length to ensure that segregation occurs along the long nucleoid axis.
<b>Perturbed nucleoid simulations</b>			Parameter values as above unless specified below. See also Materials and Methods for further details.
$k_{mB}$	Plasmid-bound mobile ParA-ATP hydrolysis (into ParA-ADP) stimulated by ParB.	$4 \text{ s}^{-1}$	Fitted, value is chosen to simulate the effect of a disordered nucleoid structure, allowing mobile ParA-ATP to diffuse past plasmids.
$S$	Short axis length of the nucleoid region where nucleoid bound ParA-ATP can polymerize.	30 nm (short), 10 nm (long)	Fitted, values are chosen to ensure a sufficient amount of mobile ParA-ATP.

doi:10.1371/journal.pcbi.1004009.t004

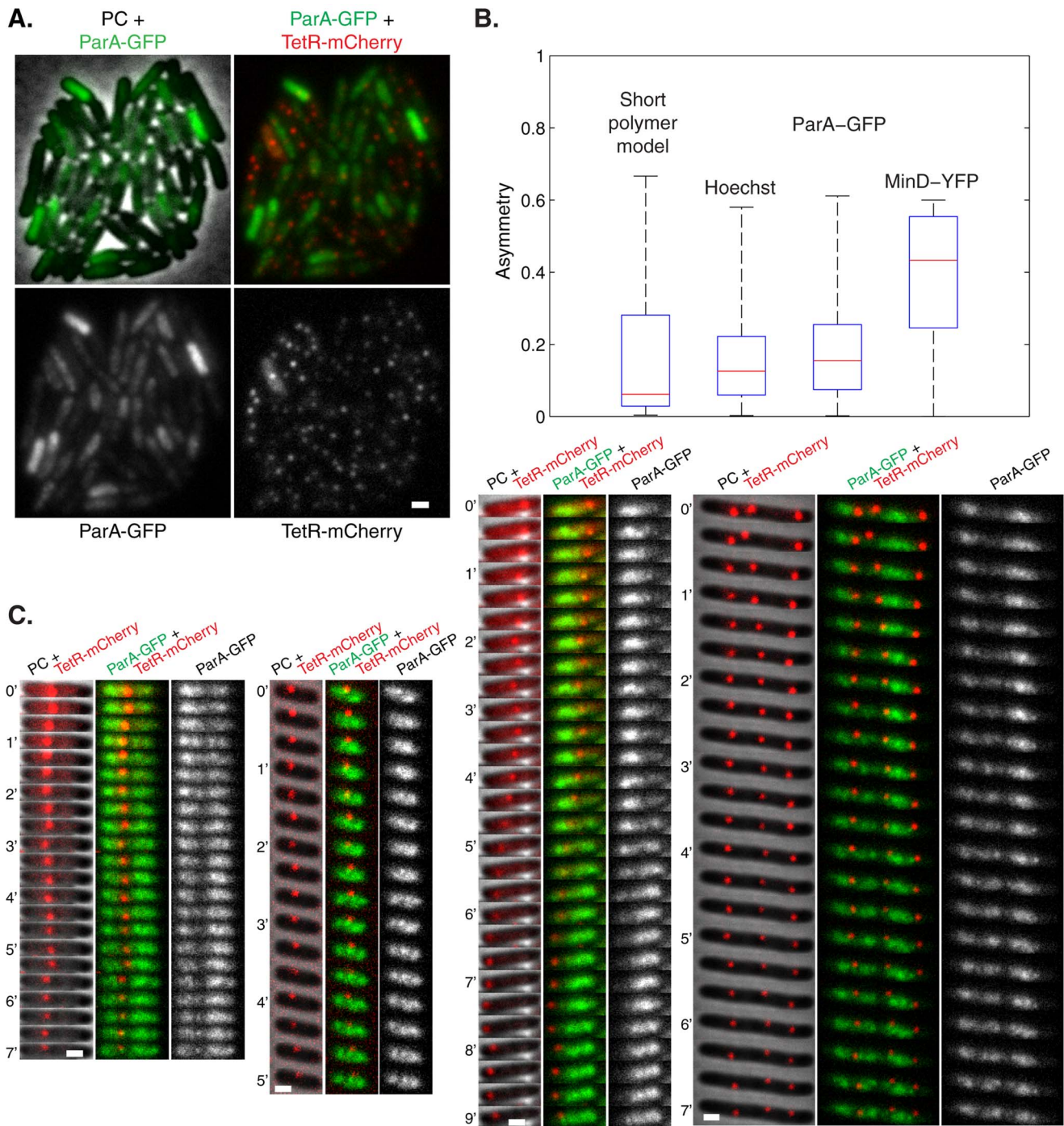
nucleoid as concluded above. However, it is unclear why ParA-ATP on either side of a plasmid would align in a coherent 1d-like structure with their ends coinciding with a plasmid. One potential explanation for this 1d-like behavior is that the ParA-ATP structures are sensitive to the overall nucleoid architecture. To test these features, we examined the localization of ParA-GFP and Hoechst signal simultaneously using optical sectioning in WT cells ( $n = 678$ ) without *par2*-carrying plasmids to prevent dynamic ParA-GFP structure disassembly. ParA-GFP intensity correlated well with the DNA stain (Fig. 6A,B, S6A, Pearson's correlation coefficient  $r_p = 0.81$ ), indicating that ParA-GFP localization was indeed dependent on the underlying nucleoid. Importantly, ParA-GFP overlaid more with Hoechst than the reverse (Fig. 6C), indicating that ParA forms structures within the nucleoid region rather than uniformly covering the nucleoid. Although the resolution of our techniques does not allow identification of potential individual ParA polymers, in many cases we did observe

extended 1d-like ParA-GFP structures on the nucleoid (Fig. 6B, S6A). Care must be taken in interpreting fluorescent localization studies due to potential artifacts, for example GFP-induced polymerization [31]. However, wild-type plasmid loss rates and plasmid foci positioning in cells expressing ParA-GFP argue against localization or polymerization artifacts in our case [9,12].

### Equal plasmid spacing is compromised in cells with a perturbed nucleoid

We reasoned that if ParA structures are reliant on the nucleoid morphology for their organization, then mutants/treatments that perturb the overall nucleoid structure should also exhibit alterations in ParA localization and therefore plasmid focus positioning (Fig. 7A). We measured plasmid focus positioning in *mukE*, *mukF* and *matP* mutant strains, as well as in cells treated with the DNA gyrase inhibitor nalidixic acid (Nal), all of which exhibit defects in nucleoid organization [32–34]. Nucleoid length

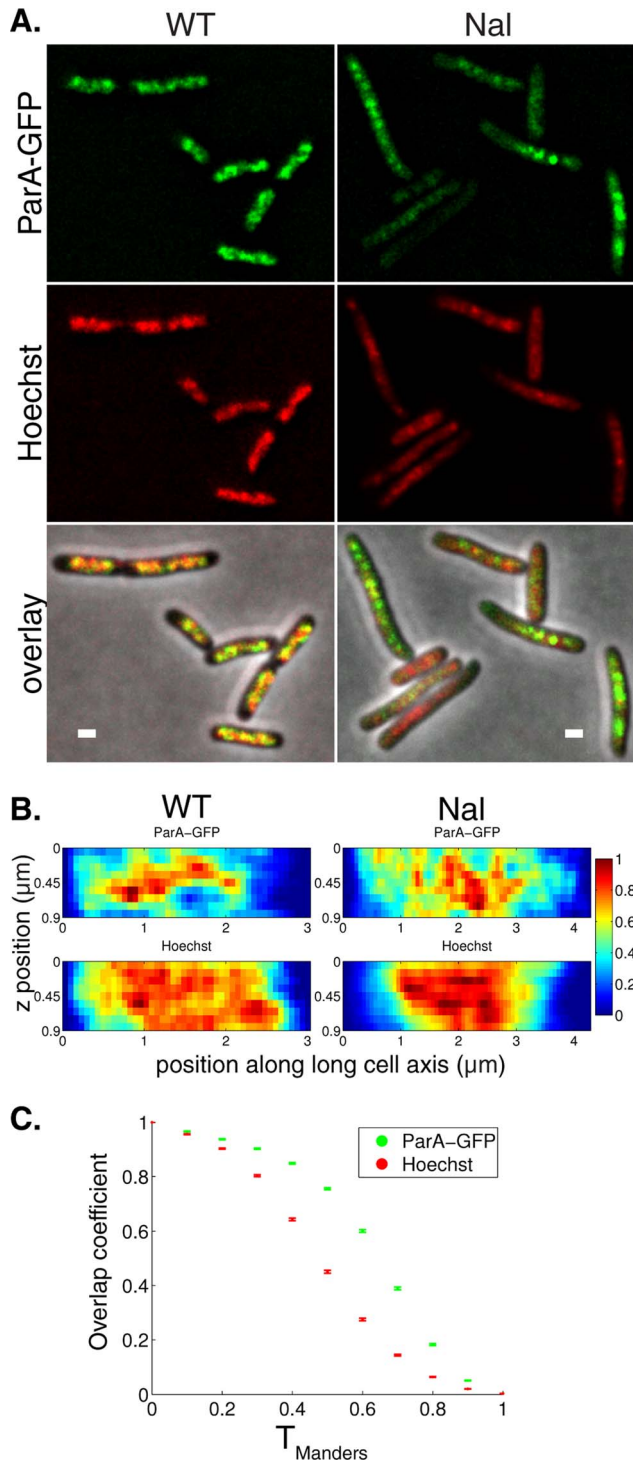




**Fig. 5. As predicted by the directed motion model, ParA-GFP distribution is relatively symmetric.** (A) Localization of plasmids and summed Z-stack of ParA-GFP distributions in a field of cells. Scale bar: 1  $\mu$ m; plasmid: pSR233 (mini-R1, *par2<sup>+</sup>*, *P<sub>lac</sub>::parA::eGFP*, *tetO120*) in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). (B) ParA-GFP ( $n = 134$ ) asymmetry measure  $|I_L - I_R| / (I_L + I_R)$  using fluorescence densities  $I_L$ ,  $I_R$  on left, right sides of a plasmid focus along long cell axis (see Materials and Methods). Comparison shown to the prediction of directed motion model with short polymers, Hoechst ( $n = 134$ ) and MinD-YFP case [27]. Box plots represent minimal, first quartile, median, third quartile and maximal values of asymmetries in all cases. (C) Time lapses showing localization of *par2<sup>+</sup>* pSR236 (mini-R1, *parC1<sup>+</sup>*, *parA<sup>+</sup>*, *parB<sup>+</sup>*, *parC2<sup>+</sup>*, *P<sub>lac</sub>::parA::eGFP*, *tetO120*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). doi:10.1371/journal.pcbi.1004009.g005

distributions were altered in all of these cases (S7A Fig.) and, consistent with our hypothesis, there was in each case a similar deterioration in the fidelity of plasmid focus positioning ( $n_p = 1, 2$  in Fig. 7B, S7B,C,  $n_p = 3, 4$  in S7B,C,D Fig.) towards a random distribution (S7E Fig.). This deterioration may not have been large

enough to detect in stability assays [35,36]. Similarly, in *E. coli mukB* mutants, perturbed plasmid positioning without compromising plasmid stability has also been observed, although for the segregation mechanism mediated by ParM [37]. The deteriorations in plasmid positioning could have resulted from other effects,



**Fig. 6. ParA forms structures within the nucleoid region.** (A) Fluorescence localization of ParA-GFP (green), Hoechst DNA stain (red) and overlay, at mid-height through cell, taken from deconvolved Z-stacks showing structures that are disrupted with 50  $\mu\text{g}/\text{ml}$  nalidixic acid treatment (Nal) compared to WT. Scale bar: 1  $\mu\text{m}$ ; plasmid: pGE230 (mini-R1, *par*, *P<sub>lac</sub>::parA::eGFP*). (B) Normalized fluorescence intensity profiles along the long cell axis for 9 in focus z heights ( $\text{dz}=0.1 \mu\text{m}$ ) resulting from deconvolved Z-stacks in representative WT and Nal-treated strains. (C) Manders overlap coefficients in WT cells (error bars: standard error of the mean,  $n=678$ ) showing the fraction of ParA-GFP fluorescence intensity that overlaps with Hoechst DNA stain when the latter is above a threshold  $T_{\text{Manders}}$  (ParA-GFP, green) and the reverse

(Hoechst, red). ParA-GFP overlaps more with Hoechst DNA stain (p-values ranging from  $10^{-12}$  to  $10^{-132}$ , see Materials and Methods) than the reverse.

doi:10.1371/journal.pcbi.1004009.g006

such as an induction of the SOS response in Nal-treated cells. However, the similarity of the altered plasmid positioning in all four cases instead suggests a common positioning defect based on nucleoid perturbation. This deterioration could also be due to an altered plasmid structure. However, at least for the case of *matP* we do not favor this hypothesis, due to the absence of MatP target sites (*matS*) on our test-plasmid.

To provide evidence that the above deterioration in plasmid positioning arose from an altered ParA distribution, we systematically examined localization of ParA-GFP and Hoechst stain simultaneously in Nal-treated cells ( $n=862$ ), which had the largest visible perturbations. We were able to quantify (Fig. 7C, S6B,  $p<10^{-149}$ , see Materials and Methods) perturbations in nucleoid structure that were detectable by eye (Fig. 6A, S6A). Moreover, visual inspection showed that the ParA-GFP distribution followed the nucleoid structure less closely than in the WT (Fig. 6A, S6A). This finding was quantitatively confirmed by a correlation coefficient of  $r_P=0.68$ , decreased from its WT value of 0.81 ( $p<10^{-34}$ ), and also by a decrease in the ParA-GFP overlap coefficient (Fig. 7C). Altogether, these findings support our hypothesis that the nucleoid provides a template for 1d-like ParA-ATP structure formation, which is partially compromised when the nucleoid structure is perturbed.

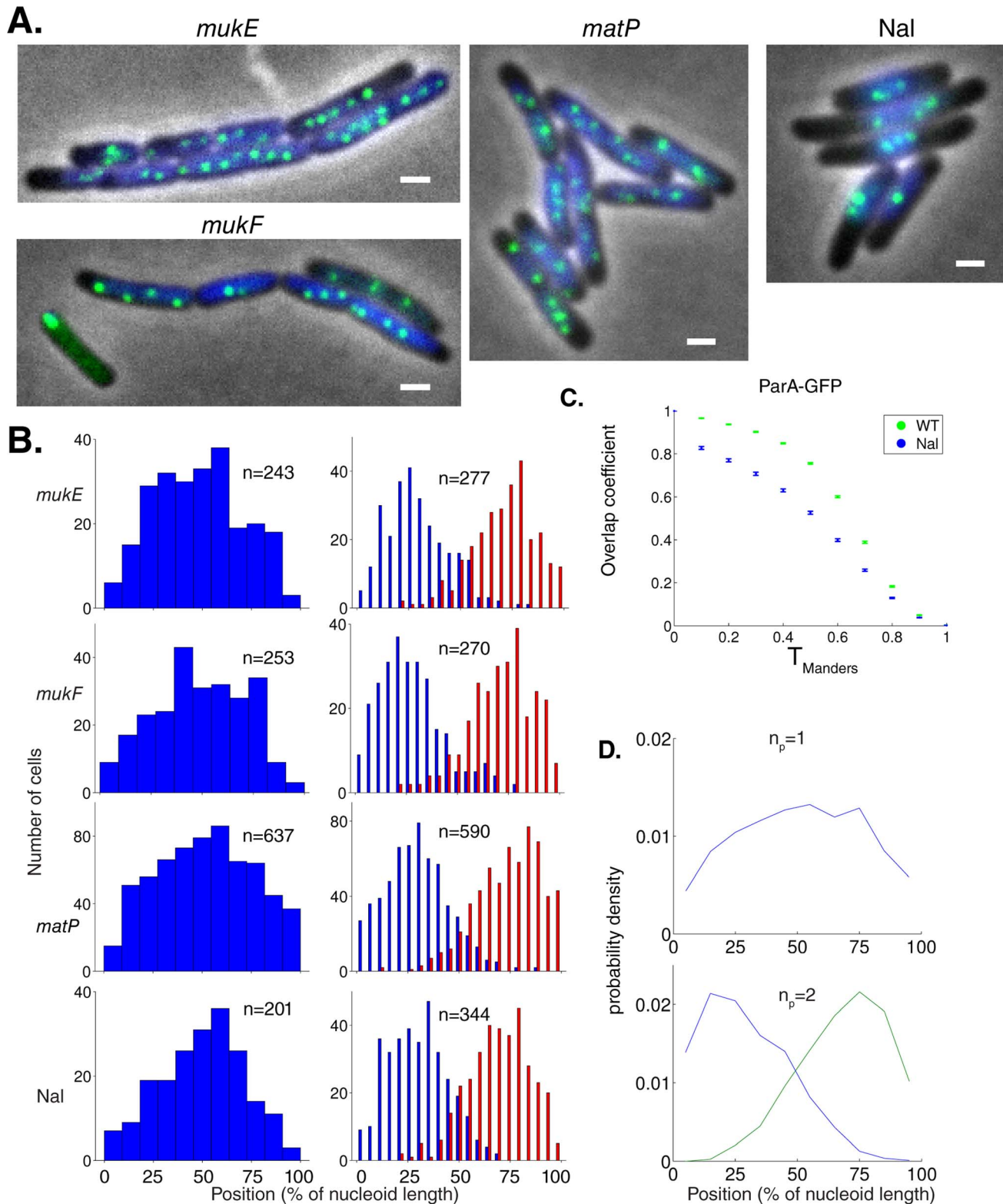
To reproduce this behavior in the directed motion model, we assumed that mobile DNA-bound ParA-ATP could now diffuse past a plasmid (see Materials and Methods for details). This could be due to the disordered nucleoid structure resulting in a deteriorated ParA-ATP structure organization, thereby allowing ParA-ATP to spatially bypass ParB-*parC* complexes and compromise the ParA concentration differences between either side of a plasmid. The directed motion model with a weak (Fig. 7D, S7F Fig.) or strong (S7F Fig.) degree of polymerization could then reproduce the observed plasmid focus distributions (Fig. 7B).

## Discussion

Stable DNA inheritance is important for the viability of essentially all organisms. In bacteria, the *parABC* genes have a major role in this process for plasmid DNA [1]. In this study, we have investigated how *E. coli* utilizes the *par2* partitioning system from plasmid pB171. We have for the first time provided a robust mechanistic explanation for how plasmids are equally spaced over the nucleoid, a process vital for the fidelity of low copy number plasmid inheritance. We propose that competing ParA structures function to direct plasmid movement over the nucleoid to equally spaced positions. This mechanism is likely relevant to other *parABC* systems that move and position sub-cellular objects.

It has previously been proposed that plasmid positioning is controlled by concentration gradients of ParA-ATP over the nucleoid, caused by plasmid-associated ParB-*parC* complexes mediating ParA-ATP hydrolysis [8,16–20]. In this so-called diffusion-ratchet mechanism [8,17,18,20], it has remained unclear whether such a mechanism could actually mediate equal plasmid spacing, and if so, which specific properties of the system were key. In particular, it was left unclear how ParA actually influenced plasmid movement [8,17,18], e.g. through immobilizing plasmids or actively directing their motion through a chemophoresis force [19,20]. Furthermore, although the diffusion-ratchet mechanism did not strictly preclude some degree of ParA polymerization, its





**Fig. 7. Nucleoid morphology disruption causes aberrant plasmid focus positioning.** (A) Fluorescence localization of ParB-GFP (green) and Hoechst DNA stain (blue) in *mukE*, *mukF*, *matP* mutants and wild-type cells treated with 50  $\mu\text{g/ml}$  nalidixic acid (Nal). Scale bar: 1  $\mu\text{m}$ ; plasmid: pFS21 (mini-R1, *parC1*<sup>+</sup>, *parA*<sup>+</sup>, *parB::sfGFP*, *parC2*<sup>+</sup>). (B) Histograms of plasmid foci positions ( $n_p = 1, 2$ ) for mutants/treatments described in (A) relative to nucleoid size. According to Kolmogorov-Smirnov tests, all distributions are broader than WT (Fig. 1C) with  $p < 10^{-2}$  except Nal  $n_p = 1$ :  $p < 0.05$ . (C) Manders overlap coefficients (error bars: standard error of the mean) of ParA-GFP comparing WT ( $n = 678$ ) and Nal-treated cells ( $n = 862$ ). Consistent with a decrease in the Pearson's correlation coefficient  $r_p$  ( $p < 10^{-38}$ ), ParA-GFP overlaps less with Hoechst in Nal-treated cells as compared to WT ( $p$ -values ranging from  $10^{-51}$  to  $10^{-144}$ ). (D) Time-averaged plasmid position distributions for directed motion model with short polymers obtained as in Fig. 2C from 124 independent simulations. Here, mobile DNA-bound ParA-ATP was now able to diffuse past a plasmid.  
doi:10.1371/journal.pcbi.1004009.g007

gradient-aspect was emphasized as opposed to polymerization [8,16–18,20], leaving open the potential importance of polymerization. To provide elucidation of these key issues, we have therefore performed a mathematical analysis, which has led to predictions that we have experimentally verified.

We found that ParA-ATP nucleoid-binding, followed by diffusion over the nucleoid, and subsequent ParB-*parC*-stimulated ParA unbinding in a 1d model, is sufficient to generate dynamic ParA-ATP concentration gradients on either side of a plasmid. We have further shown that these ParA concentrations on either side of a plasmid are only symmetric in the case of equally spaced plasmids; unequally-spaced plasmid configurations will cause the ParA gradient to be steeper on one side rather than the other. Fundamentally, this asymmetry arises from two key properties: (i) a greater space for binding of ParA on one side as opposed to the other in unequally-spaced configurations, and (ii) ParA only being returned to the cytoplasm at discrete plasmid positions occupied by ParB-*parC*. The combination of these two features leads to the ParA density being increased in larger versus smaller inter-plasmid regions and hence to asymmetric ParA concentrations in unequally spaced plasmid configurations. According to our analysis, all that is then required for equal plasmid spacing is that the plasmids have a means to preferentially move up the locally steepest ParA concentration gradient and thus locate the equally spaced configuration with symmetric, competitive ParA concentrations around each plasmid. The exact means of plasmid translocation is therefore not critical; all that is important is that such movement can occur.

With this general framework established, we then investigated which specific means of plasmid movement up a concentration gradient were possible, and which was implemented for the *par2* segregation system. We first developed a diffusion/immobilization model and found that such a model could indeed lead to plasmid movement up a ParA gradient, as the plasmid tends to become trapped in regions of higher ParA concentration. However, when we tested this model experimentally, its predictions did not verify: in particular, plasmid mobility was higher in the presence rather than the absence of *par2*, and overall free plasmid mobility was too low to allow the experimentally-observed rapid plasmid segregation following duplication events. This intrinsically low mobility agrees with earlier measurements [23,26,38] and is likely a general feature for relatively large intracellular components, given the glass-like properties of the cytoplasm [23].

We then considered active means of ParA-mediated plasmid movement. In particular, we assumed that ParA-ATP could form polymeric filaments, which could subsequently depolymerize through the action of plasmid-associated ParB-*parC*. In this case, ParA-ATP could bind to the nucleoid, diffuse and then subsequently polymerize to form gradients of ParA polymers, with the degree of polymerization influenced by the overall ParA concentration at a particular location. We found that ParA polymer models could naturally explain enhanced plasmid mobility in the presence of *par2*, as well as rapid plasmid segregation events, much more satisfactorily than the diffusion-immobilization model, regardless of whether long or short ParA polymers were formed. This finding in particular shows that our directed motion model is sufficiently general to explain equal plasmid spacing as found in various *parABC* systems with different extents of ParA polymerization [8,9,18]. In addition, we note that this mechanism does not critically depend on ParA-ATP binding to the nucleoid as a dimer. A scenario where ParA polymerizes to a certain extent cytoplasmically, and subsequently binds and

diffuses on the nucleoid before polymerizing further into immobile filaments, could also suffice.

A key aspect of our models is competition between ParA structures on either side of a plasmid to direct plasmid movement. Therefore our model predicts a comparatively symmetric ParA distribution on average, a prediction which we experimentally verified. We note here that such competition makes the system dynamics robust to alterations in ParA expression levels, since it is only the relative rather than absolute ParA levels on either side of a plasmid that are critical. This analysis potentially explains why cells with variable amounts of ParA-GFP (S1C Fig.), still possess functional segregation systems with low plasmid loss rates [9].

In the above polymer models, the movement of a plasmid is assumed to be directed by retracting ParA structures. The precise nature of this short-ranged directed motion is not specified by our analysis, and could include locally biased plasmid diffusion along a retracting polymer in a “burnt-bridge” mechanism [4] or even direct pulling [39]. This arbitrariness is a special case of our more general result that the mechanism by which a plasmid is able to move up a ParA concentration gradient is not important, only that such movement is possible. Other mechanisms of directed motion are also plausible. One possibility is that ParA-ATP does not polymerize at all, but nevertheless forms dense structures on the nucleoid with many ParA-ATP contacting a plasmid at any given time. In this variant, biased diffusion through an analog of a “burnt-bridge” mechanism is still possible. Another possibility is a DNA-relay, where directed motion is generated by the elastic dynamics of the nucleoid DNA to which ParA-ATP dimers are bound [21]. Moreover, plasmid diffusion seems not always required for directed plasmid movement. Brownian dynamics simulations based on ParB-*parC*-mediated disassembling ParA polymer bundles can both tether and pull plasmids simultaneously without the need for plasmid diffusion [39]. We propose that distinct underlying translocation mechanisms, as exemplified above, could be responsible for directed motion in different *parABC* systems and yet still attain similar equal plasmid spacing.

For our models to generate equal plasmid spacing, ParA should be organized into a 1d-like configuration along the nucleoid. If ParA were not organized in this way, it would be possible for ParA to diffuse around the sides of a plasmid without encountering the hydrolyzing effect of the ParB-*parC* complex. This would equalize the ParA concentrations on both sides even in the case of asymmetrically placed plasmids, leading to failure of the equal spacing mechanism. Potentially such ParA structures could consist of long ParA polymer bundles, or an extended region containing short ParA polymers or dimers. Importantly, in this work, we have provided experimental evidence for such ParA structure formation within the nucleoid region. Interestingly, it has been reported that the *E. coli* chromosome adopts a helical shape [28,30]. Potentially the ParA structures could be preferentially located within a “valley” in this configuration, thereby naturally generating a 1d-like appearance, even for dimers or short polymers. Consistent with these concepts, we found experimentally that plasmid positioning is compromised in nucleoid perturbed strains. ParA structures could also provide a high enough ParA concentration to ensure plasmid tethering and directed plasmid motion, whilst preventing plasmids from diffusing away from the nucleoid, a process which would compromise regular positioning. Further investigation of the exact involvement of the nucleoid in intracellular cargo positioning is therefore an important future goal.

## Materials and Methods

### Diffusion/immobilization model

On the one dimensional lattice with sites of size  $dx = 5$  nm, sites are numbered  $0 \dots (L-1)$ . Reactants are  $A_i$ : ParA-ATP at site  $i$  with number  $A[i] (\geq 0)$ ,  $P_{ji}$ : plasmids with  $j$  ParA-ATP bound to it at site  $i$  with number  $P[j][i] (\geq 0)$ ,  $A_{ADP}$ : cytoplasmic ParA-ADP with number  $A_{ADP} (\geq 0)$ ,  $A_{cyto}$ : cytoplasmic ParA-ATP with number  $A_{cyto} (\geq 0)$ . The reactions and corresponding propensities  $p_t$  are described in Table 1. Parameter values used are listed in Table 2.

We varied the exact number of ParA-ATP molecules forming a complex that are required to completely immobilize the plasmid and this variation does not alter the qualitative behavior of the system. Introduction of a low spontaneous ParA-ATP hydrolysis parameter  $k_{off}$  also does not alter the behaviour of the system. We do not keep track of the spatial positions of ParA-ADP and ParA-ATP in the cytoplasm. Instead we merely keep track of their number.

The ParA concentration is assumed to be constant throughout the cell cycle, consistent with the total ParA-GFP fluorescence as a function of cell volume when expressed from an inducible promoter (S1C Fig.). In accordance with estimates for average ParA copy numbers obtained by semi-quantitative Western blots (S1B Fig.), the ParA concentration is assumed to be 2400 ParA (dimers) per  $\mu m$  of nucleoid. Simulations start at time  $t = 0$  and run until time  $t$ , updated according to the Gillespie algorithm, exceeds a predefined time  $T$ . To simulate nucleoid growth during the cell cycle the nucleoid lattice is extended by two sites of size  $dx$  (not containing any ParA or plasmids), at one randomly chosen position along the nucleoid length. Such a growth event occurs at regular time intervals. Reaction propensities are then updated in accordance with the new state.

In Fig. 2B the nucleoid grows from  $1.5 \mu m$  to  $3 \mu m$  in  $T = 40$  min, reflecting one cell cycle. Initially a quarter of the total ParA in the system is in the cytoplasmic ParA-ADP form, 11 ParA-ATP are bound to each plasmid to ensure initial anchoring, and the rest are bound randomly to the nucleoid. In Fig. 2B the plasmid is initially located at site 0. In the simulations used to generate the histograms shown in Fig. 2C, S3A, all plasmids are initially distributed randomly across the nucleoid. At regular time intervals of 5 s the simulation state is output along with the plasmid positions to generate a time-averaged probability distribution for the plasmid positions along the long axis of the cell. In cases where the total number of plasmids ( $n_p$ ) is more than one, the plasmids are ordered and labeled  $1 \dots n_p$  according to their positions (by increasing site number) along the nucleoid. Their position is then used to generate distributions for every plasmid label  $1 \dots n_p$  for that particular overall number of plasmids  $n_p$ .

In the event of plasmid duplication at a particular site where an existing plasmid is located, a new plasmid without any bound ParA is added to the same site and the reaction propensities are updated accordingly. In case of two or more existing plasmids, one is chosen randomly for duplication. Plasmid duplication events in Fig. 2B occur at regular time intervals  $T/3$ , although the model behaves equally well with duplication at any time as it dynamically segregates the plasmids to equally spaced positions.

### Directed motion model

The nucleoid was represented as a rectangular lattice divided into square sites of sides  $dx = 5$  nm. The long axis could grow from  $1.5 \mu m$  to  $3 \mu m$  in length, while the short axis of the nucleoid lattice remained fixed. For wild-type directed motion model

simulations the short axis length was 30 nm (directed motion model with short polymers) and 25 nm (directed motion model with long polymers). Thus every site had a coordinate along the long axis (labelled  $0 \dots L-1$ ) as well as a coordinate along the short axis (labelled  $0 \dots S-1$ ). Reactants are:  $A_{mij}$ : mobile ParA-ATP at site  $(ij)$  with number  $A_{m[i][j]} (\geq 0)$ ,  $A_{ij}$ : polymeric ParA-ATP at site  $(ij)$  with number  $A[i][j]$  (0 or 1);  $P_i$ : plasmids at site  $i$  with number  $P[i] (\geq 0)$ ;  $A_{ADP}$ : cytoplasmic ParA-ADP with number  $A_{ADP} (\geq 0)$  and  $A_{cyto}$ : cytoplasmic ParA-ATP with number  $A_{cyto} (\geq 0)$ . The reactions and corresponding propensities  $p_t$  are listed in Table 3. Parameter values used are listed in Table 4.

In the perturbed-nucleoid simulations, mobile ParA-ATP can diffuse past a plasmid with 10% (short) or 100% (long) of the normal diffusion rate and the short axis length of the nucleoid is altered to 10 nm in the long polymer model. Lastly, to allow for mobile ParA-ATP to move past the plasmid without being hydrolyzed,  $k_{mB}$  is reduced 10-fold compared to its standard value.

As for the diffusion/immobilization model, the total ParA concentration was constrained to be 2400 ParA (dimers) per  $\mu m$  of nucleoid (long axis) and the total length of simulated time was  $T = 40$  min. Initially a quarter of the total number of ParA in the system was in the cytoplasmic ParA-ADP form, with the rest distributed randomly on the nucleoid in the mobile ParA-ATP form. Initial plasmid positioning, state output, plasmid position distribution generation and plasmid duplication rules were also as described previously. Nucleoid growth was implemented as described previously, with one generalization: a position along the long axis of the nucleoid was first chosen randomly. Then two nucleoid slices of 1 site (along the long axis) by  $S$  sites (along the short axis) were inserted.

### Plasmids and strains

The ParA-GFP fusion and *tetO*-TetR-mCherry plasmid labeling system were described previously [9,12]. To obtain the functional ParB-GFP fusion, the *parB* gene in the *par2* locus was replaced by *parB::sfGFP* and inserted in a mini-R1 test-plasmid. See S2 Text for more details on the strains and plasmids construction, semi-quantitative ParA western blotting and supplemental figure data analysis.

### Epifluorescence microscopy

*E. coli* strains carrying plasmids of interest (see Table S1, S2 in S2 Text for details on strains and plasmids) were grown to stationary phase while being shaken at  $37^\circ C$  in LB medium supplemented with appropriate antibiotics (30 or 50  $\mu g/ml$  ampicillin, 25 or 50  $\mu g/ml$  kanamycin, 15  $\mu g/ml$  chloramphenicol), with the exception of the *muk* strains, which were grown at  $24^\circ C$ . Cultures were diluted to an  $OD_{450}$  of 0.025 in antibiotic-free M9 minimal medium containing supplements (0.2% casamino acids, 0.2% glycerol, 1  $\mu g/ml$  thiamine, 1 mM  $MgSO_4$ , 0.1 mM  $CaCl_2$ ). Inoculated cultures were incubated until an  $OD_{450}$  of  $\approx 0.2$  was reached, typically taking 3 h.

When nalidixic acid was used to condense the nucleoids, the antibiotic was added to a growing culture at a final concentration of 50  $\mu g/ml$  two hours before imaging. Where appropriate, culture samples were mixed with Hoechst 33342 (Invitrogen) at a final concentration of 50  $\mu g/ml$  for DNA staining immediately before microscopy.

For imaging, cells were immobilized on 1.5% agarose-M9 pads mounted on microscopy slides using Gene Frames (Thermo Scientific). All microscopy experiments, unless specified otherwise (see below), were carried out using an Olympus IX71 inverted microscope with a CoolSNAP HQ EMCCD digital camera

(Photometrics, pixel size = 0.066  $\mu\text{m}$ ). A temperature-controlled incubation chamber (Applied Scientific) fitted to a Weather Station (Precision Control) kept samples at a constant 30°C. Images were acquired using SoftWoRx version 5.5.0 with a Zeiss Plan-Neofluar 100X/1.30 NA oil objective and Olympus Mercury 100 W burner (U-LH100HG) fluorescent light source. Filter set specifics are given in Table S3 in S2 Text.

### Optical sectioning of fluorescence signals from ParA-GFP and Hoechst-stained nucleoid DNA

Expression of ParA-GFP from plasmid pGE230 (mini-R1, *par*<sup>-</sup>, *P*<sub>lac</sub>::*parA::eGFP*) in *E. coli* strain KG22 or FS1 (KG22 $\Delta$ matP) was induced by adding 10  $\mu\text{M}$  IPTG to the culture medium two hours before microscopy. A 31 image Z-stack with 0.1  $\mu\text{m}$  section widths was taken for all projections (exposure times Phase Contrast (GFP channel): 0.05 s, ParA-GFP: 1.5 s, Phase Contrast (Hoechst channel): 0.1 s, Hoechst: 2 s). Image stacks were subsequently deconvolved using SoftWoRx v.5.5.0 with the following parameters: 10 iterations, medium noise reduction, conservative method.

### Measuring asymmetry in ParA-GFP distributions using optical sectioning

Expression of ParA-GFP from plasmid pSR233 (mini-R1, *par*2<sup>+</sup>, *P*<sub>lac</sub>::*parA::eGFP*, *tetO120*) in *E. coli* KG22 cells harboring pSR124 (*P*<sub>BAD</sub>::*tetR::mCherry*) was induced by adding 10  $\mu\text{M}$  IPTG to the culture medium one hour before microscopy. Samples were treated with Hoechst stain and imaged immediately thereafter. Expression of TetR-mCherry was not induced, as baseline activity of *P*<sub>BAD</sub> produced sufficient amounts of TetR-mCherry to detect foci in a single image at mid Z-height. Similarly, a single Hoechst stain image was acquired. For ParA-GFP, a 21 image Z-stack with 0.2  $\mu\text{m}$  section widths was taken (exposure times Phase Contrast: 0.1 s, TetR-mCherry 1.5 s, ParA-GFP: 1.5 s, Hoechst 0.15 s). Images were acquired using a Zeiss Axiovert 200 M inverted epifluorescence microscope with a Zeiss Plan-Neofluar 100X/1.3 NA oil objective in a temperature-controlled room at 22°C. The microscope was controlled using MetaMorph software version 7.7.5.0 (Molecular Devices, Inc.). Cells were illuminated using a Lambda LS xenon-arc lamp and images acquired using a CoolSnap HQ<sup>2</sup> EMCCD digital camera (Photometrics, pixel size = 0.0625  $\mu\text{m}$ ). Filter set specifics are given in Table S3 in S2 Text.

### Time-lapse imaging of plasmid foci movement

Plasmid foci of the *par*<sup>-</sup> mini-R1 plasmids pMH82tetO120 (*par*<sup>-</sup>, *tetO120*<sup>+</sup>) or pSR236 (*par*C1<sup>+</sup>,  $\Delta$ *parA*, *par*B<sup>+</sup>, *par*C2<sup>+</sup>, *P*<sub>lac</sub>::*parA::eGFP*, *tetO120*<sup>+</sup>) in *E. coli* strain SR1 (KG22 $\Delta$ *pcnB*) were visualised by labelling *tetO* arrays on the plasmid *in trans* with TetR-mCherry provided from the pSR124 vector (see [40] for the original method). TetR-mCherry expression was induced by adding L-arabinose to a final concentration of 0.02% to growing cultures for 15 minutes, followed by catabolite repression with 1% glucose for 10 minutes. In strains harbouring pSR236, expression of ParA-GFP was induced by the addition of 10  $\mu\text{M}$  IPTG inducer 2 h before microscopy. Time-lapse image series were acquired for different total durations/time intervals: 1 min/4 s or 15 min/30 s for pMH82tetO120 (exposure times phase contrast: 0.1 s; TetR-mCherry: 1.5 s) and 1 min/5 s or 15 min/20 s for pSR236 respectively (exposure times phase contrast: 0.1 s; TetR-mCherry: 1.5 s; ParA-GFP: 1 s). The maximum rate of image acquisition possible with our imaging system was every 4 s and 5 s (without and with ParA-GFP channel) for pMH82tetO120 and pSR236 respectively. Sample focus was maintained in the mid-cell plane

throughout the experiment using the UltimateFocus system (Applied Precision) sampling and refreshing before the acquisition of each individual frame.

### Plasmid focus positioning microscopy

*E. coli* strains KG22, FS1 (KG22 $\Delta$ matP), FS2 (KG22: *mukE::kan*) or FS3 (KG22: *mukF::kan*) harbouring pFS21 (mini-R1, *par*C1<sup>+</sup>, *par*A<sup>+</sup>, *par*B::sfGFP, *par*C2<sup>+</sup>) were grown to an OD<sub>450</sub> of 0.3. Samples were treated with Hoechst stain and imaged immediately in the mid-cell plane (exposure times ParB-GFP: 1 s, Hoechst: 0.5 s). Images of *muk* strains were acquired using a Zeiss Axiovert 200 M inverted epifluorescence microscope with a Zeiss Plan-Neofluar 100X/1.3 NA oil objective in a temperature-controlled room at 22°C. The microscope was controlled using the MetaMorph software version 7.7.5.0 (Molecular Devices, Inc.). Cells were illuminated using a Lambda LS xenon-arc lamp and images acquired using a CoolSnap HQ<sup>2</sup> EMCCD digital camera (Photometrics, pixel size = 0.0625  $\mu\text{m}$ ). Filter set specifics are given in Table S3. Other strains were imaged using both the Olympus IX71 and Zeiss Axiovert 200 M systems described above.

### Plasmid foci mobility determination

Using the MATLAB-based software suite MicrobeTracker (MT) [27], we determined *E. coli* cell outlines from phase contrast (PC) images, as well as the distribution of *tetO*-TetR-mCherry-labeled plasmids along the long axis of cells. The cell outlines were used together with the MATLAB tools spotFinderZ and spotFinderM [27] to determine *tetO*-TetR-mCherry foci positions in *par*<sup>-</sup> time-lapses of 1 min (short) or 15 min (long) in duration with images taken at intervals of 4 s or 30 s respectively. The linear *tetO*-TetR-mCherry distribution was used to control the peak detection method for false positives/negatives. For the short time-lapses we analysed cells with one or more foci, although all our results were unchanged if analysis was restricted to one focus cells to prevent potential foci labelling errors. For the long time-lapses, we only analysed cells exhibiting one focus. This was due to difficulties in distinguishing between multiple foci due to merging/splitting events, out of focus plane movement and photobleaching when acquiring images using a time interval of 30 s. These effects could have resulted in biases in the analysis due to labelling errors. We were unable to lower the time interval and simultaneously image for long time periods due to TetR-mCherry photobleaching.

At every time point the two-dimensional squared foci displacements  $r^2(\tau)$  after time lag  $\tau$  were determined. All measured displacements for the same time lag were then averaged together to obtain Mean Square Displacements (MSD)  $\langle r^2(\tau) \rangle$  with time lags from 4 s to 15 min (Fig. 3B,C,S3B). The measured plasmid displacement  $r_p(\tau)$  can report the true plasmid displacement  $r_p(\tau)$  at a resolution no greater than our measurement error, which can be up to 0.1  $\mu\text{m}$  due to microscope drift. Our measurements are also limited by a finite pixel size of 0.066  $\mu\text{m}$ . We therefore have:  $r(\tau) = r_p(\tau) + \epsilon$ , where  $\epsilon$  is the error due to both of the above effects. Squaring and averaging over many plasmid trajectories results in an MSD:  $\langle r^2(\tau) \rangle = \langle r_p^2(\tau) \rangle + \langle \epsilon^2 \rangle + \langle 2\epsilon r_p(\tau) \rangle$ . The last term vanishes due to averaging, but the second term remains and generates a small time independent value for  $\tau > 0$ . Even at short timescales of up to a minute, the MSD has a nonlinear shape, as has been reported before [26]. This is fully consistent with subdiffusive motion on these timescales. We thus expect the experimentally observed planar MSD for free particle subdiffusion in three dimensions to have the form:  $\langle r^2(\tau) \rangle = 4D\tau^\alpha + \beta$ . We

carefully measured the *par*<sup>+</sup> MSD up to 1 min with short time intervals between measurements (Fig. 3B). We performed a nonlinear least squares fit (weighted by the standard error of the mean (SEM):  $1/\text{SEM}(\tau)$ ) for  $\langle r^2(\tau) \rangle = 4D\tau^\alpha + \beta$  resulting in the values  $\alpha = 0.78 \pm 0.04$ ,  $D = 6.8 \pm 1.2 \times 10^{-4} \mu\text{m}^2\text{s}^{-\alpha}$ ,  $\beta = 6 \pm 1 \times 10^{-3} \mu\text{m}^2$  ( $R^2 = 0.99$ , p-values:  $4 \times 10^{-10}$ ,  $1 \times 10^{-4}$  and  $8 \times 10^{-4}$  respectively). On longer timescales up to 15 min (Fig. 3C), plasmid mobility also showed subdiffusive behaviour with a similar analysis giving  $\alpha = 0.78 \pm 0.05$ ,  $D = 6.2 \pm 2.1 \times 10^{-4} \mu\text{m}^2\text{s}^{-\alpha}$ ,  $\beta = 4 \pm 1 \times 10^{-2} \mu\text{m}^2$  ( $R^2 = 0.99$ , p-values:  $8 \times 10^{-15}$ ,  $8 \times 10^{-3}$  and  $2 \times 10^{-3}$  respectively). Analysing the two datasets combined (Fig. 3C) also generated consistent results, although the constant  $\beta$  was not significantly different from zero in this case:  $\alpha = 0.73 \pm 0.02$ ,  $D = 9.7 \pm 1.3 \times 10^{-4} \mu\text{m}^2\text{s}^{-\alpha}$ ,  $\beta = 1.6 \pm 2.4 \times 10^{-3} \mu\text{m}^2$ , ( $R^2 = 0.99$ , p-values:  $8 \times 10^{-15}$ ,  $8 \times 10^{-3}$  and 0.50 respectively, fit shown in Fig. 3C). Fitting  $\langle r^2(\tau) \rangle = 4D\tau^\alpha$  instead to this combined data set did not alter our estimates for  $\alpha$  and  $D$  significantly. On all observable timescales (i.e. 4 s and longer) the experimentally found *par*<sup>+</sup> MSD is bounded from above by the function  $4D_f\tau$ , with  $D_f = 10 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ . Moreover, free diffusion with diffusion constant  $D_f$  inside a box of cellular dimensions still exceeds the experimental subdiffusive mobility.

### Determining plasmid foci positions

Cell outlines and linear projections of ParB-GFP and Hoechst signal distributions along their long axis were determined as described above using MicrobeTracker (MT) [27]. ParB-GFP foci detection of snapshots was also performed using the methods described above. The positions of the half-maxima of the linear Hoechst signal distribution in every cell were then determined. We defined the nucleoid length as the length between the two half-maxima of the Hoechst stain. This analysis allowed us to determine the positions of plasmid foci with respect to the nucleoid.

### ParA asymmetry analysis

Here, we summed 6 planes that are in focus from a Z-stack of ParA-GFP fluorescence signal images ( $dz = 0.2 \mu\text{m}$ ), although the results are not different when using the ParA-GFP signal obtained from single confocal planes focused at mid-cell. Cell outlines, linear projections of ParA-GFP, *tetO*-TetR-mCherry and Hoechst stain fluorescence signal distributions, and *tetO*-TetR-mCherry foci positions were determined as described above. We confirmed that positioning of the *tetO*-TetR-mCherry foci from this dataset was similar to that measured previously [9]. In cells containing one plasmid focus ( $n = 134$ ), the ParA-GFP fluorescence signal from pole to plasmid position was summed and divided by the respective pole-to-plasmid distance. This generates two ParA-GFP fluorescence densities  $I_L$  and  $I_R$  for either side extending to the two cell poles. This allows us to compute the normalized ParA asymmetry measure  $|I_L - I_R| / |I_L + I_R|$ . Irrespective of the plasmid position, a completely uniform fluorescence distribution would give an asymmetry value of zero. On the other hand, if all the ParA-GFP was located on one side of the plasmid the asymmetry measure would be one. Using a single confocal plane focused at mid-cell, we also computed the Hoechst asymmetry measure with respect to the plasmid position in the same manner.

As shown in [27] by using the same MT software package for analysis, the MinD-YFP asymmetry measure with respect to mid-cell follows an approximate sinusoidal oscillation over time, with a cell-length-dependent oscillation amplitude. In large cells the MinD-YFP oscillations are clearest with an amplitude  $|I_L - I_R| / |I_L + I_R|$  of around 0.6. To generate an asymmetry

measure appropriate for the MinD-YFP oscillations, we sampled  $10^3$  time points  $t$  uniformly in  $[0, 2\pi]$  (which constitutes one period). We then computed for every time point  $|I_L - I_R| / |I_L + I_R| = |0.6 \times \sin(t)|$ . The resulting asymmetry distribution (Fig. 5B) therefore reflects the experimental MinD-YFP asymmetry with respect to mid-cell in large cells [27]. In this way, we can directly compare the asymmetry present in the ParA-GFP and Hoechst signal distributions with that induced by the spatiotemporal oscillations of MinD-YFP. We also generated asymmetry measures using our directed motion model. In simulation outcomes shown in Fig. 4B (directed motion model with short polymers) and S4B (directed motion model with long polymers), the plasmid position, cytoplasmic ParA-ADP, cytoplasmic ParA-ATP, nucleoid-bound mobile ParA-ATP and polymeric ParA-ATP levels on either side of the plasmid were output at regular time intervals of  $dt = 5$  s during a time period prior to plasmid duplication (first 2 min and 1.5 min of simulated time for directed motion model with short and long polymers respectively). Cytoplasmic ParA was assumed to be uniformly distributed throughout the cell (independently of the plasmid position), thus effectively only contributing to the denominator  $|I_L + I_R|$ . With this information we computed the ParA asymmetry using the same method as described for the experimental data. Results are shown in Fig. 5B (short polymers) and S5B Fig. (long polymers). It should be noted that according to both models, the ParA asymmetry remains very low once a plasmid is stably positioned at mid-cell, pushing the asymmetry distribution further towards zero over time. This is consistent with time lapses where stable equally spaced plasmid foci positioning correlates with ParA-GFP on either side of a plasmid focus (Fig. 5C and [9]).

### Three dimensional nucleoid and ParA structure analysis

To compare the extent of overlay and 3D structure of Hoechst (nucleoid DNA) stain and ParA-GFP, we first had to align the Z-stack pairs in an unbiased manner. To achieve this, one phase contrast (PC) image (at mid  $z$  height) of the Hoechst signal sections was aligned with one GFP section PC image (at the same  $z$  position) using the TurboReg ImageJ plugin (option: translation) [41], after cropping both PC images to match the output size of the deconvolved Z-stacks. Using the same translation as for the Hoechst PC image, Hoechst Z-stacks were then translated in ImageJ to align them with the ParA-GFP Z-stacks.

We determined cell outlines in MT as described above using the PC image acquired with the GFP channel and excluded cells that did not show visible ParA-GFP and Hoechst stain simultaneously. We then computed the linear distributions (for every  $z$  height) along the long cell axis for the deconvolved Hoechst and ParA-GFP Z-stacks. We next determined for the ParA-GFP and Hoechst signals separately in every cell the maximal intensity value in the whole cell ( $I_{\text{maxcell}}$ ) and the maximal values at every  $z$  height ( $I_{\text{max}}(z)$ ). To find the 9  $z$  planes from the Z-stacks ( $dz = 0.1 \mu\text{m}$ ) that are in-focus for each cell in an automated fashion, we summed  $I_{\text{max}}(z)$  over 9 consecutive  $z$  positions including a given starting plane and determined the starting plane that gave the largest associated summed value. This starting plane and its 8 consecutive planes formed the in focus plane set. We verified that this method generated the right focus planes by inspecting the chosen planes visually for several cells. This method circumvents the problem of different focus planes for cells on the same image stack as well as alignment inaccuracies in the  $z$  direction between ParA-GFP and Hoechst signals which are difficult to control for manually.

Visual comparison of the nucleoid shape between WT and nalidixic acid (Nal) treated cells (Fig. 6A,B, S6A) revealed clear

differences. In Nal-treated cells, the nucleoid signals, where present inside a cell, were more uniform along the long cell axis than in the WT (S6A,B Fig.). Shape differences were also visible in the raw Z-stacks suggesting they were not artefacts of the deconvolution method. To quantify these shape differences in an unbiased and systematic manner, we performed the following analysis (S6B Fig.).

We reasoned that a more uniform pattern would result in a profile along the long axis that resembled a first harmonic (first non-constant term of a Fourier expansion) between the nucleoid edges. Such a harmonic would not fit so well to a more spatially oscillating pattern that would arise, for example, from helical structures. Using the Hoechst stain  $I_{\max\text{cell}}$  and the  $I_{\max}(z)$  arising from the 9 relevant focus planes we determined the half-maximum intensity locations along the long cell axis closest to the cell poles  $x_L$  and  $x_R$  at every  $z$  height. At every focus plane  $z$  height we could now define the ‘first harmonic’ function defined for  $x_L \leq x \leq x_R$ :

$$H(x,z) = \frac{I_{\max}(z)}{I_{\max\text{cell}}} \left[ \frac{1}{2} + \frac{1}{2} \sin\left(\frac{\pi(x-x_L)}{x_R-x_L}\right) \right]$$

For every  $(x,z)$  we calculated the squared error  $SE(x,z)$  between the actual intensity value  $I(x,z)$  and  $H(x,z)$ :  $SE(x,z) = [I(x,z) - H(x,z)]^2$ . Lastly we summed over the SEs at every  $(x,z)$  and divided by the number of position points  $(x,z)$  to obtain a single measure of deviation  $SE_{\text{cell}}$  in a cell that is independent of the number of data points (and thus nucleoid size) and expression level variation between cells (because of normalization to  $I_{\max\text{cell}}$ ). We then performed a Wilcoxon rank sum test on the set of  $SE_{\text{cell}}$  comparing a population of WT cells with nucleoid-perturbed cells ( $n_{\text{WT}} = 678$  and  $n_{\text{Nal}} = 862$ ). Nucleoid shapes in Nal-treated cells were indeed altered ( $p < 10^{-149}$ ). Note that this method did not detect a notable shape change in *matP* cells ( $n_{\text{matP}} = 579$ ), potentially due to our techniques not being sufficiently sensitive.

To quantitate the colocalization of ParA-GFP and Hoechst signal in each cell, we also calculated, for every cell, the Pearson’s correlation coefficient  $r_P$  using all the intensity values  $I_{\text{ParA-GFP}}(x,z)$  and  $I_{\text{Hoechst}}(x,z)$  [42].

To determine the fraction of ParA-GFP intensity signal that overlaps with Hoechst signal and vice versa we computed Manders overlap coefficients [42]. This method requires a choice of threshold  $T_{\text{Manders}}$  to distinguish between positions  $(x,z)$  that are considered to contain or lack sufficient intensity signal. We therefore performed our analyses for the complete range of threshold values to show that our qualitative conclusions are insensitive to the choice of a particular  $T_{\text{Manders}}$  (Fig. 6C, 7C). Manders overlap coefficients of ParA-GFP and Hoechst were calculated as follows:

$$M_{\text{ParA-GFP}} = \frac{\sum_{x,z} I_{\text{ParA-GFP,coloc}}(x,z)}{\sum_{x,z} I_{\text{ParA-GFP}}(x,z)},$$

with

$$I_{\text{ParA-GFP,coloc}}(x,z) = \begin{cases} I_{\text{ParA-GFP}}(x,z) & \text{if } \frac{I_{\text{Hoechst}}(x,z)}{I_{\text{Hoechst}_{\max\text{cell}}}} \geq T_{\text{Manders}} \\ 0 & \text{otherwise} \end{cases}$$

Likewise the Manders overlap coefficient of Hoechst onto ParA-GFP is defined as:

$$M_{\text{Hoechst}} = \frac{\sum_{x,z} I_{\text{Hoechst,coloc}}(x,z)}{\sum_{x,z} I_{\text{Hoechst}}(x,z)},$$

with

$$I_{\text{Hoechst,coloc}}(x,z) = \begin{cases} I_{\text{Hoechst}}(x,z) & \text{if } \frac{I_{\text{ParA-GFP}}(x,z)}{I_{\text{ParA-GFP}_{\max\text{cell}}}} \geq T_{\text{Manders}} \\ 0 & \text{otherwise} \end{cases}$$

Note that taking  $T_{\text{Manders}} = 0$ , will generate an overlap coefficient of one by construction. The normalization to  $I_{\max\text{cell}}$  in determining the colocalizing positions allows the overlap coefficients to be comparable between cells.

In a small fraction of cells the alignment procedure described above did not result in proper alignment. This is clearly reflected in the  $r_P$  values being considerably lower for these cases than for the cell population mean  $r_P$  value. However, without excluding these few, possibly false negative, cases the population mean  $r_P$  value is still high (0.81 and 0.68 for WT and Nal-treated cells respectively), indicating that ParA-GFP and Hoechst signals generally correlate strongly at a population level. Poor alignment affects Manders overlap coefficients for the ParA-GFP and Hoechst signals on average equally and is not biased towards a particular strain/treatment. Therefore the observed misalignment of a small fraction of cells does not affect the qualitative conclusions that we state in this study.

Note that in *matP* cells, we did not observe any significant alteration in intensity correlation ( $r_P = 0.80$  for *matP*), nor ParA-GFP overlap coefficient, as compared to the WT. This result was expected given that we could not detect any significant nucleoid structure alteration, as described above.

## Supporting Information

**S1 Fig** *par2* protein functionality and expression levels. **(A)** Plasmid loss-frequency assay showing pFS21 stabilisation to wild-type levels by the recombinant *par2* locus encoding *parB::sfGFP*, confirming functionality of the fluorescent fusion protein. Plasmids used are pRBJ200 (*par*<sup>+</sup>, red), pFS19 (*par2*<sup>+</sup>, black) and experimental vector pFS21 (*parC1*<sup>+</sup>, *parA*<sup>+</sup>, *parB::sfGFP*, *parC2*<sup>+</sup>, green),  $n = 2$ , error bars: standard error of the mean. **(B)** Representative section of semi-quantitative Western blot used for approximating ParA molecule numbers *in vivo*. Cell lysate samples of strain KG22 carrying a mini-R1 plasmid lacking (pRBJ200) or containing *par2* (pGE2) were compared to plasmid-free KG22 cell lysate mixed with known amounts of purified His6ParA. Standard curve generated from intensity measurements from this blot has  $R^2 = 0.965$ . Band intensities were measured and quantified using the ImageQuant TL 1D Gel Analysis Software, ( $n = 3$ ). **(C)** Scatter plot of ParA-GFP total fluorescence signal in single WT cells as a function of cell volume, when expressed from an inducible promoter ( $P_{\text{lac}}$ ). The different color labels indicate the number of plasmid foci. Plasmids: pSR233 (mini-R1, *par2*<sup>+</sup>,  $P_{\text{lac}}::\text{parA}::\text{eGFP}$ , *tetO120*) and pSR124 ( $P_{\text{BAD}}::\text{tetR}::\text{mCherry}$ ). **(D)** Scatter plot of ParB-GFP total fluorescence signal in single WT cells as a function of cell volume, when expressed from its native promoter. Plasmid: pFS21 (*parC1*<sup>+</sup>, *parA*<sup>+</sup>, *parB::sfGFP*, *parC2*<sup>+</sup>); color labeling as in **(C)**. (PDF)

**S2 Fig** Plasmid foci are equally spaced over the nucleoid irrespective of nucleoid length or plasmid focus copy number. **(A)** Scatter plot of plasmid foci positions (blue, green, red, cyan) with respect to nucleoid edges (purple) and cell edges (black) for wild-type cells. Strains and plasmids used for S2 Fig. are as described in Fig. 1. **(B)** Histograms of plasmid foci positions shown in **(A)** relative to nucleoid length. (PDF)

**S3 Fig** Diffusion/immobilization model can move and maintain plasmids at equally spaced positions. **(A)** Time-averaged plasmid position distributions for diffusion/immobilization model with  $n_p = 3, 4$  on a simulated nucleoid growing from 1.5  $\mu\text{m}$  to 3  $\mu\text{m}$  in 40 min without plasmid duplication. Plasmid distributions were obtained by sampling positions every 5 s in 36 independent simulations. **(B)** Plots as in Fig. 3B except with experimental *par*<sup>-</sup> (red, green, blue) plasmid trajectories in which plasmid location is within a region of normalized Hoechst stain intensity  $I$  equal to or higher than the values indicated in the legend. The corresponding plasmid copy numbers ( $n_{\text{par}2+} = 763$ ,  $n_{\text{par}^-, I \geq 0} = 747$ ,  $n_{\text{par}^-, I \geq 0.5} = 592$ ,  $n_{\text{par}^-, I \geq 0.75} = 401$ ) indicate that a large fraction of *par*<sup>-</sup> plasmids do indeed reside in the nucleoid region; error bars: standard error of the mean. **(C)** Plots of 13 segregation events of *par2*<sup>+</sup> pSR236 (mini-R1, *parCI*<sup>+</sup>, *parA*<sup>+</sup>, *parB*<sup>+</sup>, *parC2*<sup>+</sup>, *tetO120*, *P<sub>lac</sub>::parA::eGFP*) plasmids in *E. coli* cells harboring pSR124 (*P<sub>BAD</sub>::tetR::mCherry*). Shown is the additionally segregated distance (colored lines) as a function of time, both with respect to the start of each segregation event. A segregation event is defined as two foci that are initially  $\leq 0.3 \mu\text{m}$  apart and subsequently segregate  $\geq 0.8 \mu\text{m}$  further apart within 20 s. The horizontal line (black) indicates 0.8  $\mu\text{m}$ . (PDF)

**S4 Fig** The directed motion model can equally space plasmids over the nucleoid, and is not critically dependent on the extent of ParA polymerization. **(A)** Time-averaged plasmid position distributions for directed motion model with short polymers with  $n_p = 3, 4$  plasmids on a simulated nucleoid growing from 1.5  $\mu\text{m}$  to 3  $\mu\text{m}$  in 40 min without plasmid duplication. Plasmid distributions were obtained by sampling positions every 5 s in 36 independent simulations. **(B)** Typical simulation kymograph of the directed motion model with long polymers. Long polymers extend from nucleoid ends in a growing cell, where plasmid (red) is initially directed from a nucleoid edge to mid-cell by ParA (green) filament competition. After plasmid duplication, the system dynamically self-organizes to reacquire equal plasmid spacing. **(C)** Time-averaged plasmid position distributions for directed motion model with long polymers with  $n_p = 1-4$  plasmids. Simulated nucleoid growth and plasmid distributions obtained as in **(A)**. (PDF)

**S5 Fig** Hoechst DNA stain and ParA-GFP signal asymmetry are relatively low and uncorrelated to plasmid focus positioning. **(A)** Scatter plot of ParA-GFP asymmetry measure as a function of cell length ( $n = 134$ ). **(B)** ParA asymmetry prediction from the directed motion model with long polymers. Comparison shown to experimental ParA-GFP ( $n = 134$ ), Hoechst ( $n = 134$ ) and MinD-YFP distributions [7]. **(C)** Scatter plot of ParA-GFP and Hoechst asymmetry as a function of (a single) plasmid focus position relative to cell length. (PDF)

**S6 Fig** Correlation between Hoechst and ParA-GFP distributions. **(A)** Normalized fluorescence intensity profiles along the long cell axis for 9 in focus  $z$  heights ( $dz = 0.1 \mu\text{m}$ ) resulting from deconvolved Z-stacks in representative WT and Nal-treated

strains. Many cases (representative examples shown) support the existence of linear ParA-GFP structures, although the inherent optical resolution of the imaging prohibits stronger conclusions about the presence or absence of narrow linear ParA-GFP filaments. For every cell having detectable Hoechst and ParA-GFP signals, the corresponding profiles were used for the systematic colocalization analyses. **(B)** Graphical illustration of the unbiased systematic ‘first harmonic’ analysis of deconvolved 3d Hoechst signal inside representative cells in WT and Nal-treated strains. The Hoechst (blue) profiles indicate the signal intensities (integrated over the cell width) along the long cell axis at 9 in focus  $z$  heights with corresponding ‘first harmonics’ (dotted red curves, see Materials and Methods). Fluorescence signal distributions deviate significantly more from the first harmonics in WT compared to Nal-treated cells (Wilcoxon rank sum test,  $p < 10^{-149}$ ), showing that Hoechst DNA stain distributions are perturbed in the latter. This analysis is independent of nucleoid length, which is altered in Nal-treated strains as compared to WT (S7A Fig.). (PDF)

**S7 Fig** Comparison of plasmid foci position histograms in cells with perturbed nucleoid morphology to completely randomized plasmid distributions. **(A)** Mean nucleoid length (error bars: standard error of the mean) of cells used for the plasmid positioning analysis shown in Figs. 1B, S2A and **(B,C)** in different strains: WT ( $n = 1695$ ), *mukE* ( $n = 1378$ ), *mukF* ( $n = 1555$ ), *matP* mutants ( $n = 2995$ ) and cells treated with nalidixic acid (Nal) ( $n = 1127$ ). According to unpaired  $t$  tests, all mutants and Nal show a mean differing from WT ( $p < 10^{-3}$ ). Although the average nucleoid length in *matP* mutants decreased, the average number of nucleoids per cell increased compared to WT ( $p < 10^{-41}$ ) due to a large fraction of cells exhibiting 2 nucleoids (using our half maximum criteria). This observation is consistent with the previously proposed function of MatP in preventing early segregation of duplicated Ter macrodomains. **(B)** Scatter plot of  $n_p = 1-4$  plasmid foci positions (blue, green, red, cyan) with respect to nucleoid edges (purple) and cell edges (black) for *mukE*, *mukF* mutant cells. **(C)** As in **(B)** for *matP* mutants and cells treated with 50  $\mu\text{g/ml}$  nalidixic acid (Nal). **(D)** Histograms of  $n_p = 3, 4$  plasmid foci positions shown in **(B,C)** relative to nucleoid size. **(E)** Histograms of  $10^5$  datasets for each of  $n_p = 1-4$ , where for each dataset plasmids are positioned in  $[0, 100]$  with a uniform distribution, independent from each other and consequently labeled  $1..n_p$  according to their position. This protocol induces an inherent spatial ordering. By comparing these distributions with the WT experimental data shown in Fig. 1C ( $n_p = 1, 2$ ) and S2B Fig. ( $n_p = 3, 4$ ) it is clear that the *parABC* system positions plasmid foci much more precisely, although the effect of active positioning becomes less clear as  $n_p$  increases. **(F)** Time-averaged plasmid position distributions for directed motion model with short and long polymers for  $n_p = 3-4$  (short) and  $n_p = 1-4$  (long) on simulated growing nucleoids without plasmid duplication. Results obtained from 124 independent simulations, where ParA-ATP could now diffuse past a plasmid (see Materials and Methods). (PDF)

**S1 Text** Supplementary text to section: Mathematical analysis shows that dynamic ParA concentrations can generate equal plasmid spacing. This text contains the derivation that in our mathematical model a symmetric ParA concentration implies equal plasmid spacing in case of ParB-*parC*-mediated ParA-ATP hydrolysis with any rate  $k_B$ . (PDF)



**S2 Text** Supplementary materials and methods including Table S1, S2 and S3. (DOCX)

## Acknowledgments

We thank Mark Buttner, David Guymer, Heath Murray, Seán Murray, Helmut Schiessel and Antonio Scialdone for discussions.

## References

- Gerdes K, Howard M, Szardenings F (2010) Pushing and Pulling in Prokaryotic DNA Segregation. *Cell* 141: 927–942.
- Gayathri P, Fujii T, Moller-Jensen J, van den Ent F, Namba K, et al. (2012) A Bipolar Spindle of Antiparallel ParM Filaments Drives Bacterial Plasmid Segregation. *Science* 338: 1334–1337.
- Fogel MA, Waldor MK (2006) A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev* 20: 3269–3282.
- Ptacin JL, Lee SF, Garner EC, Toro E, Eckart M, et al. (2010) A spindle-like apparatus guides bacterial chromosome segregation. *Nat Cell Biol* 12: 791–U746.
- Derome A, Hoischen C, Bussiek M, Grady R, Adamczyk M, et al. (2008) Centromere anatomy in the multidrug-resistant pathogen *Enterococcus faecium*. *Proc Natl Acad Sci USA* 105: 2151–2156.
- Roberts MAJ, Wadhams GH, Hadfield KA, Tickner S, Armitage JP (2012) ParA-like protein uses nonspecific chromosomal DNA binding to partition protein complexes. *Proc Natl Acad Sci USA* 109: 6698–6703.
- Savage DF, Afonso B, Chen AH, Silver PA (2010) Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* 327: 1258–1261.
- Vecchiarelli AG, Han YW, Tan X, Mizuuchi M, Ghirlando R, et al. (2010) ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. *Mol Microbiol* 78: 78–91.
- Ringgaard S, van Zon J, Howard M, Gerdes K (2009) Movement and equipositioning of plasmids by ParA filament disassembly. *Proc Natl Acad Sci USA* 106: 19369–19374.
- Ringgaard S, Ebersbach G, Borch J, Gerdes K (2007) Regulatory cross-talk in the double par locus of plasmid pB171E. *J Biol Chem* 282: 3134–3145.
- Ringgaard S, Lowe J, Gerdes K (2007) Centromere pairing by a plasmid-encoded type I ParB protein. *J Biol Chem* 282: 28216–28225.
- Ebersbach G, Gerdes K (2001) The double par locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. *Proc Natl Acad Sci USA* 98: 15078–15083.
- Ebersbach G, Gerdes K (2004) Bacterial mitosis: partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell. *Mol Microbiol* 52: 385–398.
- Ebersbach G, Ringgaard S, Moller-Jensen J, Wang Q, Sherratt DJ, et al. (2006) Regular cellular distribution of plasmids by oscillating and filament-forming ParA ATPase of plasmid pB171. *Mol Microbiol* 61: 1428–1442.
- Sengupta M, Nielsen HJ, Youngren B, Austin S (2010) P1 Plasmid Segregation: Accurate Redistribution by Dynamic Plasmid Pairing and Separation. *J Bacteriol* 192: 1175–1183.
- Hatano T, Niki H (2010) Partitioning of P1 plasmids by gradual distribution of the ATPase ParA. *Mol Microbiol* 78: 1182–1198.
- Hwang LC, Vecchiarelli AG, Han YW, Mizuuchi M, Harada Y, et al. (2013) ParA-mediated plasmid partition driven by protein pattern self-organization. *EMBO J* 32: 1238–1249.
- Vecchiarelli AG, Hwang LC, Mizuuchi K (2013) Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proc Natl Acad Sci USA* 110: E1390–E1397.
- Sugawara T, Kaneko K (2011) Chemophoresis as a driving force for intracellular organization: Theory and application to plasmid partitioning. *Biophys Soc Japan* 7: 77–88.
- Vecchiarelli AG, Neuman KC, Mizuuchi K (2014) A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proc Natl Acad Sci USA* 111: 4880–4885.
- Lim HC, Surovtsev IV, Beltran BG, Huang F, Bewersdorf J, et al. (2014) Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *Elife* 3: e02758.
- Slepoy A, Thompson AP, Plimpton SJ (2008) A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *J Chem Phys* 128: 205101.
- Parry BR, Surovtsev IV, Cabeen MT, O'Hem CS, Dufresne ER, et al. (2014) The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. *Cell* 156: 183–194.
- Weber SC, Spakowitz AJ, Theriot JA (2010) Bacterial Chromosomal Loci Move Subdiffusively through a Viscoelastic Cytoplasm. *Phys Rev Lett* 104: 238102.
- Weber SC, Thompson MA, Moerner WE, Spakowitz AJ, Theriot JA (2012) Analytical Tools To Distinguish the Effects of Localization Error, Confinement, and Medium Elasticity on the Velocity Autocorrelation Function. *Biophys J* 102: 2443–2450.
- Polka JK, Kollman JM, Mullins RD (2014) Accessory factors promote AlfA-dependent plasmid segregation by regulating filament nucleation, disassembly, and bundling. *Proc Natl Acad Sci USA* 111: 2176–2181.
- Slusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol Microbiol* 80: 612–627.
- Fisher JK, Bourniquel A, Witz G, Weiner B, Prentiss M, et al. (2013) Four-Dimensional Imaging of *E. coli* Nucleoid Organization and Dynamics in Living Cells. *Cell* 153: 882–895.
- Wiggins PA, Cheveralls KC, Martin JS, Lintner R, Kondev J (2010) Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc Natl Acad Sci USA* 107: 4991–4995.
- Hadizadeh Yazdi N, Guet CC, Johnson RC, Marko JF (2012) Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. *Mol Microbiol* 86: 1318–1333.
- Swulius MT, Jensen GJ (2012) The Helical MreB Cytoskeleton in *Escherichia coli* MC1000/pLE7 Is an Artifact of the N-Terminal Yellow Fluorescent Protein Tag. *J Bacteriol* 194: 6382–6386.
- Badrinarayanan A, Lesterlin C, Reyes-Lamothe R, Sherratt D (2012) The *Escherichia coli* SMC Complex, MukBEF, Shapes Nucleoid Organization Independently of DNA Replication. *J Bacteriol* 194: 4669–4676.
- Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, et al. (2008) The MatP/matS Site-Specific System Organizes the Terminus Region of the *E. coli* Chromosome into a Macrodomain. *Cell* 135: 475–485.
- Nollmann M, Crisna NJ, Arimondo PB (2007) Thirty years of *Escherichia coli* DNA gyrase: From in vivo function to single-molecule mechanism. *Biochimie* 89: 490–499.
- Ezaki B, Ogura T, Niki H, Hiraga S (1991) Partitioning of a Mini-F Plasmid into Anucleate Cells of the MukB Null Mutant. *J Bacteriol* 173: 6643–6646.
- Funnell BE, Gagnier L (1995) Partition of P1 Plasmids in *Escherichia coli* MukB Chromosomal Partition Mutants. *J Bacteriol* 177: 2381–2386.
- Tao WT, Dasgupta S, Nordstrom K (2000) Role of the mukB gene in chromosome and plasmid partition in *Escherichia coli*. *Mol Microbiol* 38: 392–400.
- Derman AI, Lim-Fong G, Pogliano J (2008) Intracellular mobility of plasmid DNA is limited by the ParA family of partitioning systems. *Mol Microbiol* 67: 935–946.
- Banigan EJ, Gelbart MA, Gitai Z, Wingreen NS, Liu AJ (2011) Filament depolymerization can explain chromosome pulling during bacterial mitosis. *PLoS Comput Biol* 7: e1002145.
- Lau IF, Filipe SR, Soballe B, Okstad OA, Barre FX, et al. (2003) Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol Microbiol* 49: 731–743.
- Thevenaz P, Ruttimann UE, Unser M (1998) A pyramid approach to subpixel registration based on intensity. *IEEE T Image Process* 7: 27–41.
- Manders EMM, Verbeek EJ, Aten JA (1993) Measurement of Colocalization of Objects in Dual-Color Confocal Images. *J Microsc-Oxford* 169: 375–382.

## Author Contributions

Conceived and designed the experiments: RI FS KG MH. Performed the experiments: FS. Analyzed the data: RI FS. Wrote the paper: RI FS KG MH. Constructed the theory: RI MH. Built the computational models: RI.



# Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation

Zhe Wu<sup>a,1</sup>, Robert Ietswaart<sup>a,b,1</sup>, Fuqian Liu<sup>a,2</sup>, Hongchun Yang<sup>a</sup>, Martin Howard<sup>a,b,3</sup>, and Caroline Dean<sup>a,3</sup>

<sup>a</sup>Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, United Kingdom; and <sup>b</sup>Computational and Systems Biology, John Innes Centre, Norwich NR4 7UH, United Kingdom

Contributed by Caroline Dean, November 17, 2015 (sent for review September 15, 2015; reviewed by Klaus Grasser and Kim Sneppen)

The basis of quantitative regulation of gene expression is still poorly understood. In *Arabidopsis thaliana*, quantitative variation in expression of *FLOWERING LOCUS C* (*FLC*) influences the timing of flowering. In ambient temperatures, *FLC* expression is quantitatively modulated by a chromatin silencing mechanism involving alternative polyadenylation of antisense transcripts. Investigation of this mechanism unexpectedly showed that RNA polymerase II (Pol II) occupancy changes at *FLC* did not reflect RNA fold changes. Mathematical modeling of these transcriptional dynamics predicted a tight coordination of transcriptional initiation and elongation. This prediction was validated by detailed measurements of total and chromatin-bound *FLC* intronic RNA, a methodology appropriate for analyzing elongation rate changes in a range of organisms. Transcription initiation was found to vary ~25-fold with elongation rate varying ~8- to 12-fold. Premature sense transcript termination contributed very little to expression differences. This quantitative variation in transcription was coincident with variation in H3K36me3 and H3K4me2 over the *FLC* gene body. We propose different chromatin states coordinately influence transcriptional initiation and elongation rates and that this coordination is likely to be a general feature of quantitative gene regulation in a chromatin context.

chromatin | *COOLAIR* | autonomous pathway | FCA | alternative polyadenylation

The influence of chromatin on transcription and cotranscriptional processing is of central importance in the regulation of gene expression (1, 2). An intensively studied example where the local chromatin state is considered to influence transcription in *Arabidopsis* is *FLOWERING LOCUS C* (*FLC*). *FLC* encodes a MADS-box transcription factor and acts as a floral repressor (3, 4). *FLC* expression is tuned by different genetic pathways: *FRIGIDA* activates *FLC* expression through a mechanism requiring Trithorax homologs, Paf1C, and SET DOMAIN GROUP 8 (SDG8), an H3K36 methyltransferase (5). *FLC* expression is repressed by the autonomous pathway and vernalization (5). Both these repressive pathways involve a group of antisense long noncoding transcripts collectively termed *COOLAIR*, which initiate immediately downstream of the poly(A) site at the 3' end of *FLC*. These antisense transcripts terminate at either proximal sites internal to the *FLC* gene, or distal sites within the *FLC* promoter (6, 7). Mutation of autonomous pathway components, including the RNA binding proteins FCA and FPA and the conserved components of the 3' processing complex FY, Cstf64 and Cstf77, leads to relative reduction in use of the proximal polyadenylation sites and increased *FLC* sense expression (reviewed in ref. 8). FCA localizes to *FLC* chromatin near the proximal poly(A) sites (9), and this together with the fact that PRP8 and CDKC2 (P-TEFb component), identified in FCA suppressor screens (10, 11), both require *COOLAIR* to repress *FLC*, supports the idea that promotion of proximal polyadenylation of *COOLAIR* is directly linked to reduced *FLC* expression. *FLOWERING LOCUS D* (FLD), an H3K4me2 demethylase, also functions in this mechanism and *fld* is the most effective suppressor of FCA function at *FLC* (9). FLD modulates H3K4me2 levels in the gene body of

*FLC*; however, how FCA functions with FLD to achieve *FLC* repression remains to be fully elucidated.

Here, we investigate how FCA and FLD transcriptionally repress *FLC* through analysis of Pol II occupancy. We use these data together with RNA measurements to parameterize an analytic mathematical model of *FLC* transcription. Model predictions are then tested through detailed measurements of intronic total and chromatin-bound RNA levels. This methodology is very appropriate for evaluating elongation rate changes in whole organisms where pulse-chase experiments are technically unfeasible. At *FLC*, we find that both FCA- and FLD-mediated repression occurs not only through reduced transcription initiation, but also through a coordinately reduced Pol II elongation rate. We propose that chromatin modifications at *FLC* induced by FCA and FLD, influenced by the antisense transcript processing, coordinately change initiation and elongation to quantitatively regulate the transcriptional output of the locus.

## Results

**RNA Fold Changes Do Not Reflect Pol II Occupancy Changes.** Measurement of steady-state spliced *FLC* and unspliced *FLC* RNA showed an increase in expression of ~20- to 25-fold between Col and *fca-9* and *fld-4* (Fig. 1A). We reasoned that, if this was caused by a 25× change in transcription initiation, a 25× increase

## Significance

The textbook view of how transcription is quantitatively regulated is through changes in transcription initiation. However, the arrangement of DNA in chromatin in eukaryotes and the frequent occurrence of noncoding transcripts add to the complexity of transcriptional regulation. Here, we explore the quantitative transcriptional regulation of *FLC*, a gene important for developmental timing in *Arabidopsis*. *FLC* expression correlates with altered antisense transcript processing and different chromatin states. Through experiments and mathematical modeling, we discover that transcription initiation and elongation are tightly coordinated and both are influenced by the chromatin state at the locus. Modulation of the chromatin environment by noncoding transcripts to coordinately influence transcription initiation and elongation could be a general mechanism to regulate quantitative transcriptional output.

Author contributions: Z.W., R.I., M.H., and C.D. designed research; Z.W., R.I., F.L., and H.Y. performed research; Z.W., R.I., M.H., and C.D. analyzed data; and Z.W., R.I., M.H., and C.D. wrote the paper.

Reviewers: K.G., University of Regensburg; and K.S., University of Copenhagen.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>1</sup>Z.W. and R.I. contributed equally to this work.

<sup>2</sup>Present address: Institute of Global Food Security, School of Biological Sciences, Queen's University, Belfast BT9 7BL, United Kingdom.

<sup>3</sup>To whom correspondence may be addressed. Email: caroline.dean@jic.ac.uk or martin.howard@jic.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518369112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518369112/-DCSupplemental).

in Pol II levels would be found at *FLC*, assuming transcript half-lives, splicing/3' processing efficiency, Pol II processivity, and elongation rates are unaffected in *fca-9* and *fld-4* (Fig. 1B). However, both total Pol II and productively elongating Pol II (Ser2-P) showed relatively small changes (2–3 $\times$ ) across *FLC* in the different genotypes (Fig. 1C and D, and Fig. S1A and B). We ruled out a number of technical issues with Pol II ChIP that could have led to an underestimation of Pol II occupancy. First, measurements on a highly expressed gene (*ACT7*) and a Pol IV/V transcribed region (*IGN5*) showed that a wide dynamic range (>1,000 $\times$  by comparing levels at *ACT7* to *IGN5*) could be detected in the Pol II ChIP assay (Fig. 1C and D). Pol II levels at *FLC* were well above background at *IGN5* (Fig. 1C and D, and Fig. S1). Second, specific dilutions of *FLC* chromatin, without changing the overall amount of chromatin, showed rough linearity between the Pol II ChIP signal and the Pol II concentration at *FLC* (Fig. S2). Third, cell-specific *FLC* expression variation is also highly unlikely to underlie this difference in RNA and Pol II up-regulation, as both assays use whole plant seedlings and thus reflect population averages. Based on these observations, we conclude that FCA/FLD-mediated changes in *FLC* transcription are unlikely to occur solely through changes in transcription initiation.

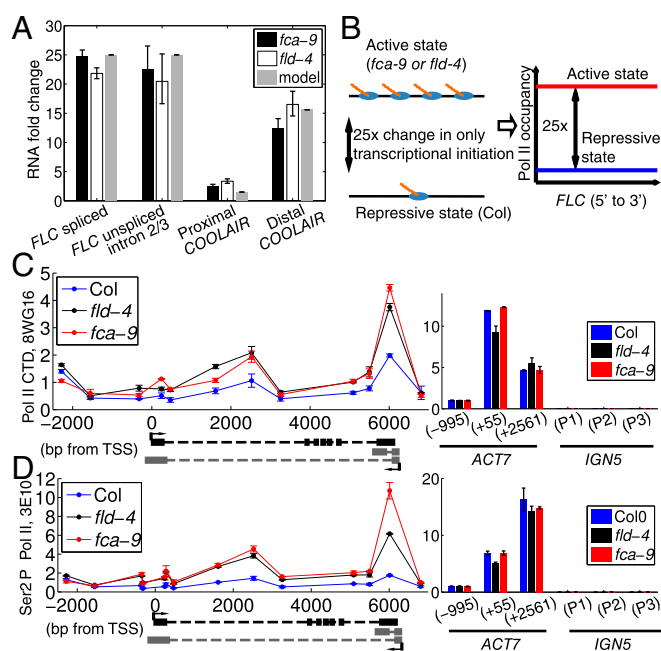
**FLC Transcriptional Dynamics Can Be Explained by Coordination of Initiation and Elongation.** To further understand how FCA- and FLD-mediated *FLC* repression occurs at a transcriptional level, we developed an analytical mathematical model of the transcriptional dynamics at *FLC* by incorporating sense *FLC* and *COOLAIR*

initiation, elongation, and termination (Fig. 2A; see [Supporting Information](#) for complete description). The experimental data described above were used as model inputs. This strategy enabled us to assign parameter values for key processes during transcription (e.g., initiation and elongation). Pol II levels reflect a density that can be described mathematically as a ratio of the initiation rate ( $F$ ) over the elongation rate ( $v$ ) (12). Because our ChIP signal is not strand specific, we summed the sense and antisense model Pol II levels to generate a model total Pol II profile along *FLC* (Fig. 2B). The small increase of Pol II ChIP signal in the transcriptionally active *fca-9* and *fld-4* mutants (Fig. 1C and D, and Fig. S1) is explained by the model through a coordinated increase in initiation and elongation rates (Fig. 2B and C). The model also reproduced the *FLC* spliced, unspliced, and *COOLAIR* fold up-regulation in *fca-9* and *fld-4* (Fig. 1A), where a 25 $\times$  fold increase in sense Pol II initiation required an 8–12 $\times$  fold faster rate of elongation to quantitatively fit the Pol II occupancy increase (Fig. 2D). Elevated Pol II levels at the 3' of *FLC* resulted from sense termination and proximal antisense transcription (Fig. 2A–D). Our model does not take into account transcriptional interference (TI) between sense *FLC* and *COOLAIR* (Discussion). Using an experimentally determined value for the termination rate  $1/50\text{ s}^{-1}$  (13), absolute elongation rates could be inferred from the model, yielding 0.2–0.4 kb/min (Col) and 1.8–3.6 kb/min (*fca-9* and *fld-4*). These correspond well to values found in other organisms (14–17). The excellent fit of the experimental data strongly supports a model where *FLC* transcriptional dynamics are governed by coordinated changes in initiation and elongation.

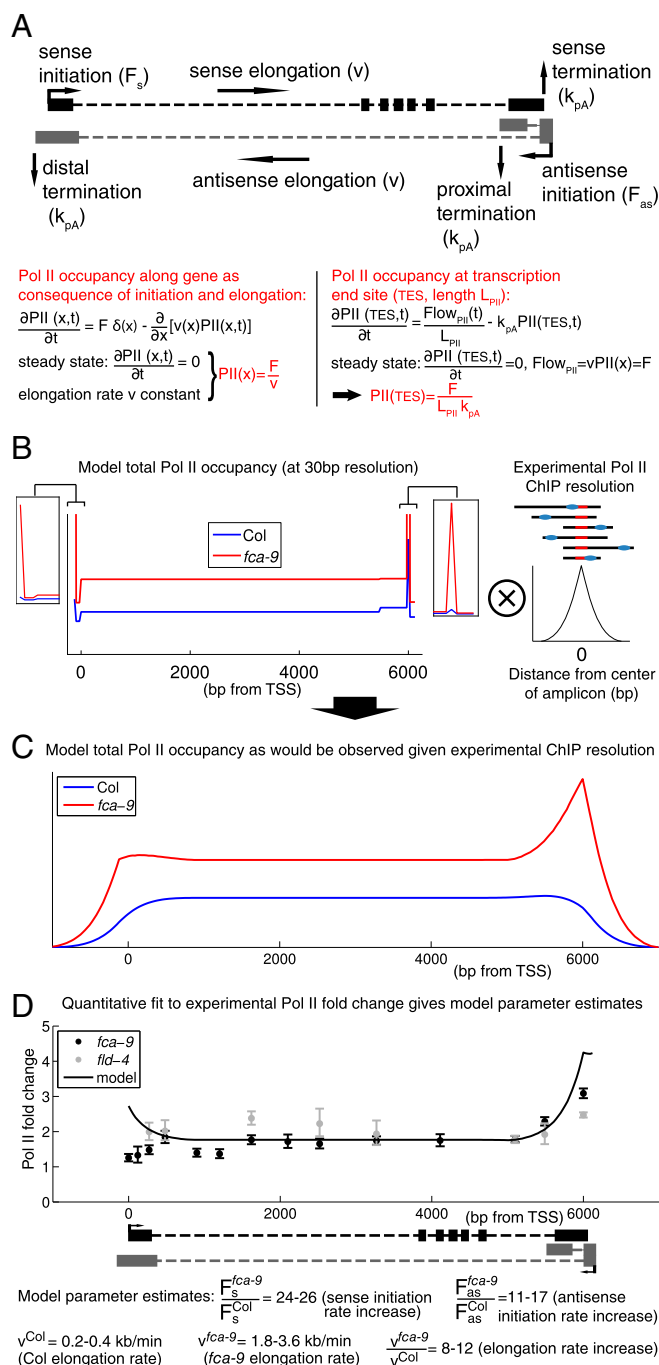
#### Cotranscriptional Splicing, Combined with Coordinated Initiation and Elongation, Generate Distinctive Patterns of RNA Up-Regulation Along *FLC* Intron1.

We next tested the predicted coordinate increase in initiation and elongation rates experimentally. Measurement of elongation rates on a subset of highly expressed, long mammalian genes (>50 kb) has been achieved using GRO-seq (14). This technique involves inhibition of elongation and then release and relies on rapid removal of an inhibitor that is difficult in whole organisms (15, 16). We tried an alternative approach via generation of an *FLC-MS2* fusion (13), but this was not expressed at a sufficiently high level to be useful. To overcome these limitations, we used our theoretical model to make specific predictions with regards to intronic *FLC* RNA production, which we then tested experimentally. If introns are spliced cotranscriptionally once Pol II has reached the 3' end of the intron, then nascent RNA from the 5' end of the intron resides on the chromatin longer than that from the 3' end. This generates a nascent RNA profile along an intron with declining levels from the 5' to 3' end (17, 18). An analytic mathematical analysis (Fig. 3A and [Supporting Information](#)) predicts that the ratio of Pol II initiation ( $F$ ) over the elongation rate ( $v$ ) determines the slope of the nascent intronic RNA levels between the 5' to 3' ends, whereas the initiation rate over the splicing rate ( $k_s$ ) determines the levels of completely transcribed, unspliced introns (Fig. 3A). This analysis indicates that nascent RNA levels close to the intron 3' end will be mostly determined by the ratio of the initiation rate to the splicing rate, and independent of the elongation rate. Away from the 3' end of the intron, transcripts emerging from Pol II still transcribing the intron will also contribute to nascent RNA levels, and hence the ratio of the initiation rate to the elongation rate will also be important (Fig. 3A). Taking into account both increased initiation and elongation rates in the *fca-9* mutant compared with Col (Fig. 3B), this analysis enabled us to predict a spatially varying fold up-regulation of nascent RNA along *FLC* intron1 (Fig. 3B).

We tested this key model prediction by measuring the chromatin-bound RNA profile at *FLC* (Fig. 3C and Fig. S4). Comparing *fca-9* to Col, the chromatin-bound fold up-regulation inside exon1 was much larger than at the exon1–intron1 junction (Fig. S4A and G), suggesting that splicing of intron1 does occur mostly cotranscriptionally. In the first kilobase of intron1, as predicted by the



**Fig. 1.** Large increases in RNA are associated with small changes in Pol II occupancy. (A) RNA fold up-regulation in *fca-9* and *fld-4* mutants compared with Col: spliced and unspliced *FLC* (~25 $\times$ ), proximal (~2 $\times$ ) and distal *COOLAIR* (~13 $\times$ ). The model values are the fits to the experimental data. Experimental values are mean  $\pm$  SEM from three to six independent samples. (B) Schematic illustration of a scenario where transcription initiation is the only difference between Col and *fca-9*, so that a 25 $\times$  fold change in Pol II occupancy should be observed as illustrated on the Right. (C and D) ChIP experiments assaying Pol II occupancy across *FLC* using the antibodies anti CTD 8WG16 (C) and anti Ser<sup>2</sup>-P CTD 3E10 (D). The bar charts at the Bottom indicate Pol II levels at various control genes. Three overlapping primer pairs are used to measure *IGN5* expression (P1–P3). Values are mean  $\pm$  SEM from two independent samples, with data presented as the ratio of Pol II at *FLC*/input at *FLC* to Pol II at *ACT7* (–995)/input at *ACT7* (–995).



**Fig. 2.** Small changes in Pol II occupancy can be explained by coordinated changes in transcription initiation and elongation. (A) Schematic of *FLC* locus and outline of the mathematical model for *FLC* transcription (details in [Supporting Information](#)). Black boxes indicate sense exons; gray boxes indicate proximal (*Upper*) and distal (*Lower*) antisense exons. (B) Total (sum of sense and antisense) model Pol II levels in Col and *fca-9* across *FLC*. The *fld-4* mutant model results are identical to *fca-9*. Shown on the *Right* is a schematic of the convolution process with experimental Pol II ChIP fragment size distribution (shown in [Fig. S3](#)). (C) Total Pol II levels in Col and *fca-9* across *FLC* from the model convolved with experimental Pol II ChIP fragment size distribution. (D) Experimental and model Pol II fold-up-regulation. Experimental values are mean  $\pm$  SEM from two to five independent samples, including data shown in [Fig. 1 C and D](#), and [Fig. S1](#). Model fold changes are ratio of profiles shown in C.

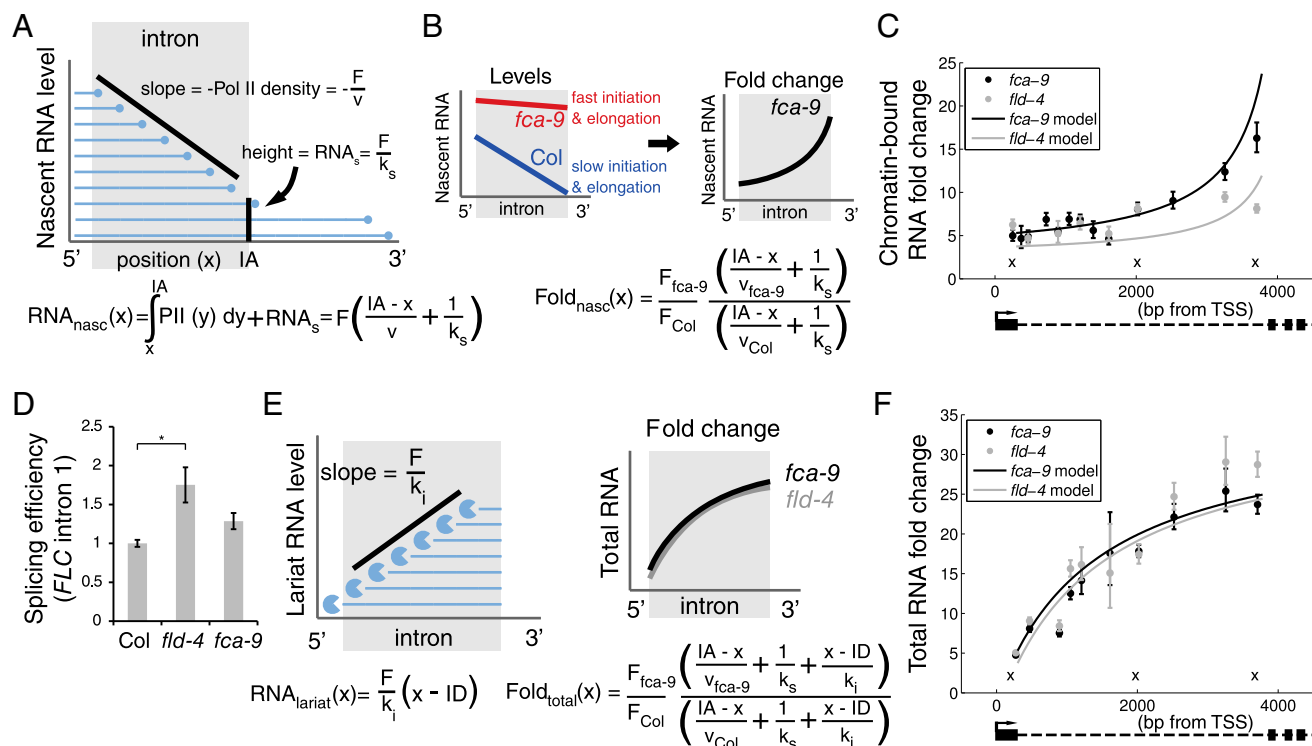
model, there was only a small fold increase in *fca-9* compared with Col ([Fig. 3C](#) and [Fig. S4A](#)). This is due to the dependence on the ratios of the initiation and elongation rates and their coordinated increases in *fca-9* ([Fig. 3B](#)). By contrast, the fold up-regulation was much larger close to the intron acceptor site in *fca-9*. This is in agreement with the model, where we used the experimentally determined splicing rate of  $1/100 \text{ s}^{-1}$  (17) for both Col and *fca-9*, with other parameters determined from our prior fitting to the Pol II ChIP data ([Supporting Information](#)). Importantly, the chromatin-bound RNA profile along intron1 is not flat, which is what would be predicted without changes to the elongation rates between *fca-9* and Col.

We also fitted the model to the chromatin-bound RNA data directly using nonlinear regression ( $R^2 = 0.89$ ,  $F$  statistic:  $P = 3 \times 10^{-14}$ ). This procedure also led to the conclusion that significant elongation rate changes [fold =  $9.8 \pm 3.8$  (mean  $\pm$  SEM),  $P = 0.03$ ] are required to explain the profile ([Supporting Information](#)). Importantly, this method does not rely on the specific values of splicing and elongation rates and is independent of Pol II ChIP data, and thus provides additional evidence for the elongation rate changes.

Interestingly, we observed less increase in fold up-regulation toward the 3' end of intron1 in *fld-4* compared with *fca-9* ([Fig. 3C](#) and [Fig. S4A](#)). Given the fold change close to an intron acceptor site is more sensitive to splicing rather than elongation rate changes ([Fig. 3B](#)), we examined whether a splicing rate change specific to *fld-4* could explain its differential fold up-regulation pattern from *fca-9* ([Materials and Methods](#) and [Supporting Information](#)). Indeed, we found that we could fit the *fld-4* profile in our model by incorporating a twofold faster splicing rate ( $1/50 \text{ s}^{-1}$ ) in *fld-4* ([Fig. 3C](#)), while keeping all other parameters unchanged. We further verified this model prediction of an increased splicing rate in *fld-4* by measuring the splicing efficiency of *FLC* intron1. As predicted, the efficiency was increased 1.8-fold in *fld-4* ([Fig. 3D](#)), but not significantly altered in *fca-9* ( $P = 0.1$ , two-sided unpaired  $t$  test). A simple alternative model with unchanged splicing and elongation rates between Col and *fld-4* would produce a constant chromatin-bound RNA fold change across intron1. That would be consistent with the chromatin-bound RNA dataset in isolation ([Fig. 3C](#)) but implies a change in the initiation rates of approximately sevenfold ([Supporting Information](#)), which is inconsistent with our earlier spliced and unspliced *FLC* RNA fold changes ([Fig. 1A](#)).

To further support these conclusions, we investigated the total intronic RNA profile ([Fig. 3E](#) and [F](#), and [Fig. S4](#)). Such measurements include intron lariat degradation intermediates, which are present in the total but not chromatin-bound RNA fraction ([Fig. 3E](#)) (17). Assuming that lariat degradation occurs from 5' to 3', lariat RNA at the 3' generally exists for longer than that at the 5'. This generates a lariat RNA profile with increasing levels from the 5' to 3' end ([Fig. 3E](#)). Importantly, incorporating this lariat population into the total intronic RNA fold up-regulation between *fca-9* and Col, without altering the model parameterization that explained the Pol II and chromatin-bound RNA, produced a predicted profile that is qualitatively different to that found for the chromatin-bound RNA ([Fig. 3B](#) and [E](#)). This prediction was also validated experimentally ([Fig. 3F](#)). Compared with the chromatin-bound RNA profile, there was a significantly larger fold increase in the first 2 kb of the total intronic RNA profile ( $P = 8 \times 10^{-7}$  and  $4 \times 10^{-7}$  for *fca-9* and *fld-4*, respectively; two-sided Welch's  $t$  test) ([Fig. 3C](#) and [F](#), and [Fig. S4A](#) and [B](#)). In the model, we could generate such a profile, by solely incorporating 5' to 3' intron lariat degradation with a rate of 1.5 bp/s (19), in line with experimentally determined intron half-lives (17). Potential additional presence of 3' to 5' degradation (19) with a rate up to 1 bp/s did not alter our conclusions ([Supporting Information](#)). The profiles for total intronic RNA look very similar between *fca-9* and *fld-4* ([Fig. 3F](#)), in contrast to the chromatin-bound data ([Fig. 3C](#)). This similarity is because





**Fig. 3.** Combination of increased initiation and elongation, with cotranscriptional splicing and lariat degradation, leads to distinct RNA profiles along *FLC* intron1. (A) Schematic indicating intronic nascent RNA,  $RNA_{nasc}$  (blue lines), arising from Pol II (blue circles) elongating through the intron and from unspliced  $RNA_s$  with full-length intron. Once Pol II has passed the intron acceptor site (IA), splicing can occur. Initiation, elongation, and splicing rates are  $F$ ,  $v$ , and  $k_s$ , respectively. Analytic expression for  $RNA_{nasc}$  is shown below. (B) Schematic (Left) indicating model profiles of nascent RNA along *FLC* intron1 in *fca-9* and Col. Between *fca-9* and Col,  $F$  and  $v$  are coordinately increased, but with the same  $k_s$ . This generates a characteristic pattern of intronic nascent RNA fold changes between *fca-9* and Col (Right) with analytic expression shown. (C) Modeled and experimentally measured chromatin-bound RNA fold changes along *FLC* intron1. The lower increase toward the 3' end in *fld-4* is due to increased splicing rate as shown experimentally in **D**. Crosses indicate positions where data are from three different, overlapping primer sets that each show similar results (Fig. S4). (D) Estimate of *FLC* intron1 splicing efficiency (intron cleavage rate) in *fld-4* and *fca-9*, normalized to the level in Col. Values are mean  $\pm$  SEM from three independent samples. Asterisks indicate statistical significance: for all of the figures in this study,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ , two-sided unpaired  $t$  test, unless specified otherwise. (E) Schematic showing effect of 5' to 3' intronic RNA degradation on lariat RNA levels ( $RNA_{lariat}$ ). Full-length lariat RNA results from splicing and is degraded with rate  $k_i$ ; ID: intron donor site. These degradation intermediates, together with the nascent RNA described in **A**, make up total intronic RNA. Fold up-regulation then generates the characteristic profiles shown. Analytic expressions for  $RNA_{lariat}$  and total intronic RNA fold changes are shown. (F) Modeled and experimentally measured total RNA fold changes along *FLC* intron1. (C and F) Experimental values are mean  $\pm$  SEM from at least three independent samples. Absolute levels are shown in Fig. S4.

the lariat RNA effectively extends the half-life of intronic RNA and therefore reduces the effect of the differential splicing rates between *fca-9* and *fld-4* (Fig. 3F). Taken together, our total and chromatin-bound intronic RNA profiles provide strong evidence that repression of *FLC* involves a coordinated change of both the initiation and elongation rates. Moreover, the methods we developed here can be used to infer elongation rate changes in whole organisms where pulse-chase experiments are not feasible.

#### Sense Premature Termination Contributes Little to *FLC* Repression.

Previous reports have linked the elongation rate to either Pol II processivity (20) or early termination (21). In these scenarios, Pol II would terminate prematurely as a result of slow elongation. Our previous analysis did not require any such premature termination. Moreover, at an intuitive level, premature termination should lead to declining levels of Pol II from 5' to 3' in the repressed case (Col) (Fig. S5A and Supporting Information). However, we found no evidence for this in our Pol II ChIP assay (Fig. 1C and D, and Fig. S1) and no short transcripts had been detected by Northern blot using an *FLC* intron1 probe (22). These findings suggest that premature termination contributes little to *FLC* repression. To further confirm this conclusion, we undertook 3'-RACE to map transcripts ending within the promoter-proximal region of *FLC*. We could detect polyadenylated transcripts that terminated within *FLC* intron1.

These transcripts all contained *FLC* exon1 and were mostly alternatively spliced with the same donor site but with a different acceptor site, compared with the conventional *FLC* intron1 (Fig. S5B). By monitoring the alternatively spliced intron associated with premature termination, we found these transcripts are of lower abundance than unspliced intron1 in Col, *fca-9*, and 35S::FCA (Fig. S5E). Therefore, sense premature termination occurs only occasionally at *FLC* and is not a major contributor to *FLC* repression.

Cotranscriptional decay of nascent transcripts by 5' to 3' exonucleases has also been proposed to influence transcriptional output (23, 24). In such a scenario, the degradation of RNA should also lead Pol II to terminate prematurely, and therefore to declining levels of Pol II from 5' to 3' in the repressed state (Col), which is again inconsistent with our Pol II ChIP data. In addition, we analyzed *FLC* expression in mutants defective for these functions (*xrn2-1*, *xrn3-3*) (25) in *Arabidopsis* and found no increase in *FLC* nascent or fully spliced *FLC* RNA levels (Fig. S6). Therefore, such a decay pathway is unlikely to play a major role in determining the overall transcriptional dynamics at *FLC*.

**FLD Alters the Local Chromatin State to Influence Transcriptional Output via Coordinated Changes in Initiation and Elongation.** We therefore continued with our investigation of coordinated initiation and elongation rates by FCA/FLD-mediated changes in chromatin



gene body Pol II levels, and Pol II elongation rates found in yeast and mammalian cells (14, 29).

How Pol II initiation and elongation are coordinated is still unclear. In *Escherichia coli*, newly initiated RNA polymerases can facilitate elongation of the leading polymerase (30). Such a mechanism is unlikely to be the case at *FLC*, because *FLC* is not highly expressed even in its active state (compared with *Actin*). Elongation is likely influenced by Pol II CTD modifications and the chromatin state (31, 32), both directly through nucleosome turn-over dynamics and indirectly via differential recruitment of elongation factors. In *Arabidopsis*, elongation factor TFIIS is required for elongation of many genes but a *tfiis* mutant does not show changed *FLC* expression (10, 33, 34). However, *FLC* expression is particularly sensitive to reduced amounts of the histone chaperone FACT (35), so it will be interesting to test whether FACT is required for the fast elongation observed in *fca-9* and the coordination mechanism. We have found here that FLD recruitment, changed H3K4me2, and the resulting changes in H3K36me3 at *FLC* are likely important for this coordination. Our analysis of SDG8 suggests that H3K36me3 is essential to maintain both a fast initiation and elongation rate at *FLC* (Fig. 5B). We therefore propose that changed histone modifications actively influence *FLC* regulation and are not just a reflection of transcription.

Our results raise the question whether there is a general need to coordinate transcription initiation and elongation. Control of gene expression may necessitate such coordination as, for instance, a slow elongation rate relative to initiation would cause an accumulation of Pol II at the promoter that would limit the number of additional Pol II molecules that can initiate through occlusion (36). Such a limit

might become even more stringent due to bursty initiation or Pol II pausing/backtracking during elongation (37). Furthermore, antisense transcription might induce a limit on initiation rates to prevent the occurrence of TI (38). However, 5' pausing of Pol II is not a feature at *FLC* (as shown by the absence of a 5' peak in Pol II ChIP), arguing against occlusion effects. The expression of sense and antisense is positively correlated at *FLC*, arguing against a major role for TI. Instead, we suggest that altered elongation rates reinforce selection of different antisense isoforms, which can then recruit different chromatin regulators to the gene, thereby modulating coordinated transcription initiation and elongation (Fig. 5B). An important question now is to understand how far the lessons from *FLC* reflect regulation mechanisms both genome- and organism-wide. Coordination between initiation and elongation could generally enhance transcription efficiency, potentially to minimize transcription-associated genome instability (39). Modulation of the deposition of different histone modifiers by noncoding transcripts may be a general mechanism to coordinately affect Pol II initiation and elongation and thus quantitatively modulate transcriptional output.

## Materials and Methods

Experimental procedures and mathematical modeling can be found in [Supporting Information](#).

**ACKNOWLEDGMENTS.** We thank Hervé Vaucheret for providing *xrn* seeds and Robert Sablowski for comments on the manuscript. We thank C.D. and M.H. group members for discussions. This work was supported by BBSRC Grant BB/K007203/1 (to M.H. and C.D.); BBSRC Institute Strategic Program GRO (BB/J004588/1); and BBSRC studentship, VSBfonds Scholarship, and Prins Bernhard Cultuurfonds Scholarship (to R.I.).

- Smolle M, Workman JL (2013) Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta* 1829(1):84–97.
- Selth LA, Sigurdsson S, Sveistrup JQ (2010) Transcript elongation by RNA polymerase II. *Annu Rev Biochem* 79:271–293.
- Sheldon CC, et al. (1999) The *FLF* MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* 11(3):445–458.
- Michaels SD, Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11(5):949–956.
- Crevillén P, Dean C (2011) Regulation of the floral repressor gene *FLC*: The complexity of transcription in a chromatin context. *Curr Opin Plant Biol* 14(1):38–44.
- Liu F, Marquardt S, Lister C, Swiezewski S, Dean C (2010) Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. *Science* 327(5961):94–97.
- Horniyk C, Terzi LC, Simpson GG (2010) The *spen* family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev Cell* 18(2):203–213.
- Ietswaart R, Wu Z, Dean C (2012) Flowering time control: Another window to the connection between antisense RNA and chromatin. *Trends Genet* 28(9):445–453.
- Liu F, et al. (2007) The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. *Mol Cell* 28(3):398–407.
- Marquardt S, et al. (2014) Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription. *Mol Cell* 54(1):156–165.
- Wang ZW, Wu Z, Raitskin O, Sun Q, Dean C (2014) Antisense-mediated *FLC* transcriptional repression requires the P-TEFb transcription elongation factor. *Proc Natl Acad Sci USA* 111(20):7468–7473.
- Ehrensberger AH, Kelly GP, Sveistrup JQ (2013) Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* 154(4):713–715.
- Brody Y, et al. (2011) The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol* 9(1):e1000573.
- Danko CG, et al. (2013) Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* 50(2):212–222.
- Fuchs G, et al. (2014) 4sUDRB-seq: Measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* 15(5):R69.
- Singh J, Padgett RA (2009) Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* 16(11):1128–1133.
- Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15(3):163–175.
- Ameur A, et al. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* 18(12):1435–1440.
- Hesselberth JR (2013) Lives that introns lead after splicing. *Wiley Interdiscip Rev RNA* 4(6):677–691.
- Mason PB, Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* 17(6):831–840.
- Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S (2013) Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. *Mol Cell* 49(1):55–66.
- Cheng Y, Kato N, Wang W, Li J, Chen X (2003) Two RNA binding proteins, HEN4 and HUA1, act in the processing of *AGAMOUS* pre-mRNA in *Arabidopsis thaliana*. *Dev Cell* 4(1):53–66.
- Brannan K, et al. (2012) mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Mol Cell* 46(3):311–324.
- Jimeno-González S, Haaning LL, Malagon F, Jensen TH (2010) The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II. *Mol Cell* 37(4):580–587.
- Gy I, et al. (2007) *Arabidopsis* FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *Plant Cell* 19(11):3451–3461.
- Ko JH, et al. (2010) Growth habit determination by the balance of histone methylation activities in *Arabidopsis*. *EMBO J* 29(18):3208–3215.
- Kim SY, et al. (2005) Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyl transferase. *Plant Cell* 17(12):3301–3310.
- Yang H, Howard M, Dean C (2014) Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis FLC*. *Curr Biol* 24(15):1793–1797.
- Mayer A, et al. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17(10):1272–1278.
- Epshtein V, Nudler E (2003) Cooperation between RNA polymerase molecules in transcription elongation. *Science* 300(5620):801–805.
- Jonkers I, Kwak H, Lis JT (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3:e02407.
- Weber CM, Ramachandran S, Henikoff S (2014) Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* 53(5):819–830.
- Van Lijsebettens M, Grasser KD (2014) Transcript elongation factors: Shaping transcriptomes after transcript initiation. *Trends Plant Sci* 19(11):717–726.
- Dolata J, et al. (2015) NTR1 is required for transcription elongation checkpoints at alternative exons in *Arabidopsis*. *EMBO J* 34(4):544–558.
- Lolas IB, et al. (2010) The transcript elongation factor FACT affects *Arabidopsis* vegetative and reproductive development and genetically interacts with HUB1/2. *Plant J* 61(4):686–697.
- Core LJ, et al. (2012) Defining the status of RNA polymerase at promoters. *Cell Reports* 2(4):1025–1035.
- Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469(7330):368–373.
- Hobson DJ, Wei W, Steinmetz LM, Sveistrup JQ (2012) RNA polymerase II collision interrupts convergent transcription. *Mol Cell* 48(3):365–374.
- Saponaro M, et al. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157(5):1037–1049.
- Wuarin J, Schibler U (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: Evidence for cotranscriptional splicing. *Mol Cell Biol* 14(11):7219–7225.
- Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499(7458):360–363.
- Tippmann SC, et al. (2012) Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol Syst Biol* 8:593.
- Gray JM, et al. (2014) SnapShot-Seq: A method for extracting genome-wide, in vivo mRNA dynamics from a single total RNA sample. *PLoS One* 9(2):e89673.